A Hybrid Swarm Intelligence Algorithm and Machine Learning Approach Based on Feature Selection to Predict Global Temperature

By

KOMEIL SHARAFI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science in Environmental Science

Supervisor:

Dr. Mohamed Tawhid

Department of Mathematics and Statistics

Faculty of Science, Thompson Rivers University

December, 2024

THOMPSON RIVERS UNIVERSITY

ABSTRACT

Global warming is one of the most significant issues in our planet, mainly driven by increasing greenhouse gas emissions and unsustainable human activities. This research introduces a new hybrid swarm intelligence algorithm integrating machine learning techniques to predict global temperature changes. Feature selection is a complex and crucial process, especially in the high-dimensional nature of climate-related issues involving multiple contributing factors. We combine five machine learning regressors with eleven nature-inspired and swarm intelligence optimization algorithms for feature selection, resulting in a total of fifty-five models. These models are evaluated using various performance indicators across multiple datasets. We apply the rank average from the Friedman test to rank and compare their performance. The Arithmetic Optimization Algorithm (AOA), which is the best individual algorithm in the initial experiments and comparisons, is hybridized with a recently introduced swarm intelligence algorithm, the Coati Optimization Algorithm (COA), to develop our proposed hybrid algorithm, named hybrid AOA-COA. This novel combination shows improved performance and efficiency through feature selection optimization, outperforming the eleven individual optimization techniques used separately. It achieved a 13% improvement in average mean squared error (MSE) across all datasets and regressors compared to the best individual algorithm, the Arithmetic Optimization Algorithm (AOA).

Key Words: Machine learning; Natural-inspired algorithms; Swarm intelligence algorithms; Feature selection; Hybridization; Global temperature.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Mohamed Tawhid, for his invaluable guidance, support, and encouragement throughout this journey. His expertise, patience, and insightful feedback have been instrumental in shaping this research and bringing it to fruition. I am truly thankful for the opportunity to learn and grow under his mentorship.

I would also like to extend my sincere appreciation to the members of my thesis committee, Dr. Peter Tsigaris and Dr. Mark Paetkau for their time, constructive criticism, and suggestions, which greatly enriched my work.

To my beloved parents, I am forever indebted for your unwavering support; thank you for your endless encouragement and for instilling in me the values of hard work and perseverance.

To my dear sister, your belief in me has been a constant source of motivation.

To my wonderful wife, Maryam, thank you for standing by my side through every challenge, for your understanding, love, and patience. Your strength and unwavering support have made this journey possible.

Lastly, to my dear, beautiful daughter, Mariana, you are my joy and inspiration. Your smiles and laughter have been the light in my darkest days, reminding me of the beauty and importance of family. This work is dedicated to you.

Thank you all for being part of this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Global Warming and Climate Change

The term global warming refers to the long-term increase in the Earth's average surface temperature, primarily due to human activities such as the burning of fossil fuels, deforestation, and industrial processes. This phenomenon has been accelerating since the late 19th century, particularly from the 1970s onward, due to a sharp increase in the concentration of greenhouse gases in the atmosphere [Karl et al., 2009]. These gases include carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and sulfur hexafluoride (SF6), which absorb infrared radiation and trap heat within the Earth's atmosphere, causing a warming effect. The Intergovernmental Panel on Climate Change (IPCC) reported that from 2011 to 2020, the global surface temperature increased by approximately 1.1°C above pre-industrial levels, which refers to the period from 1850 to 1900 [IPCC, 2023].

Climate change refers to broader changes in the Earth's climate system, including shifts in weather patterns, increased frequency and intensity of extreme weather events, and alterations in ecosystem functions. The greenhouse effect, caused by gases like carbon dioxide (CO2) and methane (CH4) trapping heat within the atmosphere, is the primary driver of these changes. Global warming, a specific aspect of climate change, is primarily driven by human activities such as the burning of fossil fuels, deforestation, and industrial processes. The increase in greenhouse gas concentrations in the atmosphere has led to the warming of the Earth's surface, atmosphere, and oceans [on Climate Change, 2021].

Human-induced climate change directly results from unsustainable energy production, land-use practices, and consumption patterns, which have significantly escalated greenhouse gas emissions. It is estimated that the current warming rate is approximately 0.2°C per decade [Karl et al., 2009]. This sustained increase in temperature is particularly concerning, as it has wide-ranging consequences not only for the atmosphere but also for the oceans, land surfaces, and ecosystems worldwide.

The latest IPCC reports show that human activities have contributed to climate change, resulting in more frequent and severe weather events like hurricanes, droughts, wildfires, and floods. These events are not isolated and have substantially damaged ecosystems, human health, and infrastructure [IPCC, 2023]. Additionally, dangerous effects such as hurricanes, tornadoes, tsunamis, volcanic eruptions, droughts, and famines have become more common, severely disrupting people's lives and leading to mass displacement and migration due to climate change [Reddy, 2021].

Impacts on Human Health and Migration: Climatic conditions have always influenced the species living in various habitats and ecosystems as they grow and evolve [Sen et al., 2023]. As climate change accelerates, its impacts on human health are becoming increasingly apparent. Rising temperatures contribute to a host of health issues, including respiratory conditions like asthma, allergies, and skin diseases. These health effects are exacerbated by poor air quality and the spread of diseases that are more prevalent in warmer climates, leading to significant public health challenges. Moreover, climate change is expected to cause mass migrations both within and across national borders, as people are displaced by extreme weather events, rising sea levels, and food insecurity. This presents severe challenges to food security and the socio-political stability of many regions [on Climate Change, 2015].

Role of Urbanization and Energy Consumption: Urbanization is a global phenomenon that contributes significantly to the rise in greenhouse gas emissions. As urban areas grow, so do energy demands, which are often met through the consumption of fossil fuels; this leads to an increase in CO2 emissions and other greenhouse gases, exacerbating the greenhouse effect. Studies have shown that urbanization is closely linked to increased incomes, which are, in turn, associated with higher energy consumption and more significant contributions to global warming. As more people migrate to urban areas, the demand for energy-intensive goods and services increases, further intensifying the emissions problem. [Kumar, 2018].

Climate Models and Future Predictions: Climate models are utilized to better understand and project the future impacts of climate change. These models are equations based on the fundamental laws of physics, chemistry, and biology, simulating the interactions between the atmosphere, oceans, land surfaces, and ice. These models are critical in estimating future changes in temperature, precipitation patterns, and the frequency of extreme weather events under various greenhouse gas emission scenarios. The IPCC relies on these models to project future warming and assess the potential outcomes of different climate mitigation strategies [IPCC, 2023].

Mitigation and Adaptation: Since human activities have triggered this dangerous phenomenon of global warming, it is crucial to take immediate and significant action to mitigate its impacts. Mitigation involves reducing the sources of greenhouse gases, such as transitioning to renewable energy, while adaptation involves adjusting societal infrastructure to minimize the risks associated with climate change. Sustainable development, changes in consumption patterns, and investment in climate-resilient infrastructure are some of the primary strategies to address the challenges posed by global warming [Holmstrom et al., 2016].

1.2 Research Question

Our general research question was, 'How can machine learning techniques be used more effectively by applying swarm intelligence algorithms for feature selection to build a reliable predictive model for global temperature?'

To address the research question, we implement five machine learning regressors that have demonstrated strong performance in previous studies. These regressors are combined with eleven metaheuristic optimization algorithms for feature selection, resulting in a total of 55 different models. We assess the performance of these models using ten datasets, which were collected and merged from various public sources to evaluate their effectiveness in different scenarios.

Afterward, we compare the performance of the models using average ranks derived from the Friedman test to identify the best feature selection optimization algorithm. Finally, we combine the top-performing feature selector with a recently introduced algorithm to create a hybrid swarm intelligence algorithm. Overall, this new algorithm surpasses all other algorithms evaluated in this study across all datasets, metrics, and regressors.

Chapter 2

Literature Review

This chapter delves into the foundational concepts and related studies on machine learning, metaheuristic algorithms, and nature-inspired techniques for feature selection. It highlights the role of machine learning regressors and swarm intelligence algorithms in solving optimization problems related to climate data. The main takeaway is understanding the background and prior studies that justify the hybrid approach proposed later in the thesis.

2.1 Machine Learning

The two most prominent technologies for visualizations and predictions have always been data science and machine learning [Mishra et al., 2018]. These technologies enable us to make precise temperature predictions and assess the current state of global warming [Mishra et al., 2018]. Machine learning techniques, which are grounded in learning theory, enable the generalization of pattern recognition to new data that is not part of the observed sample [Vapnik, 1999].

The collection of statistical techniques for examining patterns, discovering connections, and building predictive models from data sets is what machine learning is all about [Zheng, 2018]. It is a rapidly developing and useful instrument for understanding complex networks, recognizing information within large datasets, and producing predictions for the future [Koc and Acar, 2021]. , and employing machine learning algorithms offers a viable solution to this issue [Sen et al., 2023].

The use of machine learning in addressing environmental conservation issues has become more widespread and has shown promising outcomes [Chen and Chau, 2016]. Utilizing machine learning methods is an advantageous resource for identifying and forecasting climate-associated concerns [Koc and Acar, 2021]. New approaches are needed to delve into the intricacies of climate change and devise adaptation and mitigation policies that are more streamlined and impactful. Machine learning (ML) methods have become increasingly popular in various domains, climate change included, owing to technological advancements [Ladi et al., 2022].

2.2 Introduction to Regression

Regression analysis is a statistical method used to model and analyze the relationships between a dependent variable and one or more independent variables. It helps in understanding how the typical value of the dependent variable changes when any one of the independent variables is varied, while the others are held constant. This technique is widely used for forecasting, hypothesis testing, and modeling causal relationships [Draper and Smith, 1998].

2.3 Metaheuristic Algorithms

Metaheuristic algorithms are types of computational intelligence methods that are commonly applied to solve complex optimization problems [Abdel-Basset et al., 2018]. Metaheuristic refers to higher-level heuristics devised to solve a variety of optimization problems [Dokeroglu et al., 2019]. Lately, numerous metaheuristic algorithms have been effectively utilized to tackle unsolvable problems. The attractiveness of employing these algorithms to address intricate problems lies in their ability to achieve the best or optimal solutions even for high-dimensional issues in a short period of time [Dokeroglu et al., 2019].

2.4 Nature-Inspired Algorithms (NIA)

Growing body of research in the last few decades suggests that nature is a major source of inspiration for creating intelligent systems and solving complex issues. For example, animals are able to efficiently compete for food, territories, and mates because of highly tuned organs and talents that have evolved as a result of evolutionary pressure. Evolution serves as a mechanism for optimizing the parameter settings of some of these organs and capabilities, which can be thought of as refined optimization algorithms.

[Zang et al., 2010].

Nature-inspired computing involves drawing inspiration from natural processes observed in nature. This approach has led to the development of algorithms known as Nature-Inspired Algorithms (NIA), which fall under the umbrella of computational intelligence. NIAs are primarily divided into evolutionary algorithms and swarm intelligence-based algorithms. Evolutionary algorithms are modelled after the evolutionary behavior found in natural systems. Swarm intelligence (SI) based algorithms, also known as swarm optimization techniques, aim to solve specific problems by imitating the collective behavior of natural swarms. In addition to these two categories, NIAs have various other classifications based on their sources of inspiration [Agarwal and Mehta, 2014].

2.5 Swarm Intelligence Algorithms (SI)

Swarm intelligence-based approaches have been extensively utilized for addressing real-world challenges across various fields. These methods imitate the intelligent behaviour observed in populations of species such as fish, birds, ants, and bees and these species' social dynamics are studied to carry out complex tasks in nature. By modelling populations of agents that can organize, interact, and learn from each other through the sharing of information based on individual experiences, swarm intelligence techniques effectively solve problems [Mandal and Thakur, 2024]. Swarm intelligence is inspired by the collective behavior of social insect colonies and animal behavior and includes various algorithms that facilitate effective problem-solving through cooperation and self-organization. These algorithms replicate the collaboration and information exchange observed in natural swarms, helping to achieve global optimization. [Guerra et al., 2023].

In general, metaheuristic algorithms belong to the top category, with natural-inspired algorithms being a subset of that, while swarm intelligence algorithms are a specialized subset of natural-inspired ones.

2.6 Feature Selection

Feature selection involves handling features that are inappropriate, irrelevant, or unnecessary and this process extracts the most suitable features from the datasets [Liu and Yu, 2005]. Feature selection techniques can help streamline data pre-processing by reducing data efficiently, which is beneficial for developing precise data models [Jović et al., 2015]. In classification, clustering, and regression tasks, feature selection is often utilized[Jović et al., 2015]. The fundamental idea behind any method for addressing the feature selection issue is to identify a subset of the initial features [Tawhid and Ibrahim, 2019].

Feature selection primarily concentrates on eliminating irrelevant and redundant features from the original feature set and aims to choose a small subset of relevant features [Brezočnik et al., 2018]. A feature selection issue with n features has 2 to the power of n potential solutions (or feature subsets). As each additional feature is included, the complexity doubles. As a result, a dataset with 1000 features per instance has 2^{1000} possible feature subsets [Brezočnik et al., 2018]. So, it is a complex and crucial process in machine learning and definitely needs optimization.

2.7 Related Studies

Several studies have explored the application of machine learning techniques to predict global temperature and climate patterns. One study compared the effectiveness of various machine learning techniques such as linear regression, lasso, support vector regression, and random forest. The objective was to develop a sophisticated model to validate global warming trends and identify contributing factors. The research concluded that random forest performed better than other techniques in developing accurate climate models, showing higher predictive accuracy for global temperature data [Zheng, 2018]. In another study, titled ”Climate Change Forecasting using Machine Learning Algorithms,” different machine learning algorithms were evaluated, and it was found that the most effective models were the Random Forest Regressor and the XGBoost Regressor [Vishvakarma, 2022]. These models showed significant potential in predicting long-term climate trends and global warming patterns. Such predictions are critical for multiple sectors, including agriculture, energy production, and disaster management, as they help anticipate and mitigate the effects of climate change. Furthermore, linear regression was frequently employed in these studies, offering a reliable and interpretable method for forecasting climate data, with high accuracy in predicting temperature trends [Reddy et al., 2021].

Furthermore, swarm intelligence algorithms have been effectively employed for feature selection in machine learning applications. In a research titled ”Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems,” A new hybrid binary version of the bat and improved particle swarm optimization algorithm was introduced to address feature selection problems. The study's results demonstrated the efficiency of the hybrid swarm optimization algorithm in exploring the feature space for optimal feature combinations [Tawhid and Dsouza, 2018]. Another study, "Swarm Intelligence Algorithms-Based Machine Learning Framework for Medical Diagnosis: A Comprehensive Review", provides a comprehensive review of swarm intelligence (SI) algorithms and their integration with machine learning (ML) for medical diagnosis. It discusses various SI algorithms, their applications in feature selection, and how their combination with ML enhances diagnostic accuracy. While focused on medical data, the methodologies are applicable to other domains, including climate data analysis [Houssein et al., 2022]. Also "Enhanced Salp Swarm Algorithm for Feature Selection" proposes an enhanced version of the Salp Swarm Algorithm tailored for feature selection challenges in machine learning. The authors demonstrate the algorithm's effectiveness in selecting optimal feature subsets, which can be crucial for accurate predictions in various applications, including climate modeling [Bacanin et al., 2021].

The limitations of existing machine learning and swarm intelligence approaches for climate prediction are significant and multifaceted. Basic machine learning models are often hindered by their susceptibility to overfitting and inefficiency when applied to high-dimensional environmental data. This issue becomes more pronounced given the complexity and interdependence of variables such as greenhouse gas concentrations, aerosol levels, and global temperature anomalies. While swarm intelligence algorithms offer promising solutions for feature selection, they are frequently limited by their tendency for premature convergence and require extensive parameter tuning to adapt to complex, multimodal optimization landscapes [Dehghani et al., 2023, Abualigah et al., 2021a]. The current literature does not adequately address the integration of machine learning techniques with swarm intelligence algorithms in a way that balances global exploration and local exploitation—two critical components for achieving robust and reliable climate predictions.

In the domain of feature selection methods, additional gaps are evident. Many of the existing techniques do not effectively mitigate the presence of redundant and irrelevant features in high-dimensional datasets. This inefficiency often results in increased computational costs and suboptimal performance of predictive models. While feature selection is acknowledged as crucial in machine learning [Guyon and Elisseeff, 2003, Liu and Yu, 2005], there remains a lack of hybridized approaches specifically designed to handle the intricacies of climate datasets. These datasets typically involve diverse variables, such as atmospheric gas concentrations, solar irradiance, and volcanic aerosol data, which demand tailored solutions for optimal feature extraction and model generalization.

The primary objective of this research is to address the challenge of optimizing feature selection for global temperature prediction models, with a particular focus on improving accuracy and computational efficiency. This is achieved by developing a hybrid approach that leverages the strengths of machine learning regressors and swarm intelligence algorithms to handle the high-dimensional nature of environmental data.

The overarching goal is to enhance the reliability of global temperature predictions by improving feature selection processes. By reducing the dimensionality of the input space while preserving the most relevant information, the proposed methodology aims to increase prediction accuracy and computational efficiency. This research also seeks to develop scalable solutions that are robust against the complexity of environmental datasets, ensuring their applicability to diverse climate scenarios.

This study is guided by key research questions and hypotheses. The primary research question examines how hybrid swarm intelligence algorithms can improve feature selection for climate prediction models. It further investigates which combinations of machine learning regressors and optimization algorithms yield the most accurate and computationally efficient results. The study hypothesizes that integrating global search capabilities, as seen in the Arithmetic Optimization Algorithm (AOA), with the local search strengths of the Coati Optimization Algorithm (COA) will outperform existing individual algorithms. Furthermore, it is hypothesized that optimized feature selection will significantly enhance prediction accuracy and reduce computational costs in global temperature modeling.

As a consequence, the previous machine learning regressors that showed the best performance in earlier studies and various natural-inspired and swarm intelligence algorithms were selected to find the best performance. Our objective is to evaluate the effectiveness of different swarm intelligence algorithms when paired with different machine learning models, including Linear regression, Random Forest regressor, K-Neighbor regressor, XGBoost regressor, and Support Vector regressor. Additionally, we introduce a hybrid swarm intelligence algorithm, which outperforms other examined algorithms in terms of prediction accuracy and computational efficiency.

Chapter 3

Methodology

This chapter outlines the research design, detailing data sources, preprocessing steps, and the machine learning regressors employed. It describes how various nature-inspired optimization algorithms are used for feature selection and evaluated with performance metrics. The key takeaway is the systematic approach to model development and the framework for comparing 55 models across 10 datasets.

The term ”global temperature” refers to the mean temperature of the Earth's atmosphere and surface, calculated by averaging temperature readings collected from various locations worldwide over a specified time period. This metric is widely used in climate science to monitor and analyze trends in Earth's climate system, particularly changes associated with global warming and climate variability.

The calculation of average global temperature involves two major steps: data collection and statistical averaging. Data is gathered from ground-based weather stations, ocean buoys, and satellite measurements, which collectively represent the temperature of the Earth's surface (land and oceans) and lower atmosphere. To account for variations in geographical distribution and data coverage, weighted averaging techniques are applied, where regions with sparse data are assigned proportional weights to ensure balanced global representation [Hansen et al., 2010].

This measure is critical for understanding the Earth's energy balance and assessing the impacts of human activities on the climate. Increasing trends in the average global temperature over recent decades have been attributed to rising concentrations of greenhouse gases, such as carbon dioxide, methane, and nitrous oxide, due to industrialization and deforestation [on Climate Change, 2021]. Monitoring this indicator is a key component of global climate models and is used to predict future climate scenarios, assess mitigation efforts, and guide international climate policy.

3.1 Data Source and Description

The data in the datasets come from various sources, including NASA, the National Oceanic and Atmospheric Administration (NOAA), the United States Environmental Protection Agency (EPA), the Berkeley Earth organization, Kaggle and Our World in Data platforms, and the University of East Anglia. In our research, we utilize ten datasets described in the following sections. After each dataset description, we illustrate the correlation between the independent variables (features) and the target variable (temperature) for that dataset.

In Dataset 1, we have these variables: Year (1983-2008), Month, MEI, CO2, CH4, N2O, CFC-11, CFC-12, TSI, Aerosols, and Temp (Temperature). Here we have the data description (description of variables) from the reference of the dataset:
Temp is the variation in degrees Celsius between the average global temperature during that timeframe and a reference point. This information is sourced from the Climatic Research Unit at the University of East Anglia.
Also atmospheric levels of carbon dioxide (CO2), nitrous oxide (N2O), methane (CH4), trichlorofluoromethane (CCl3F, also known as CFC-11), and dichlorodifluoromethane (CCl2F2, commonly called CFC-12) are reported. This information is sourced from the Global Monitoring Division of ESRL/NOAA.
Aerosols in the dataset shows the mean stratospheric aerosol optical depth at 550 nm. This variable is linked to volcanoes, as volcanic eruptions result in new particles being added to the atmosphere, which affect how much of the sun's energy is reflected back into space. This data is from the Godard Institute for Space Studies at NASA.
TSI refers to the total solar irradiance (TSI) measured in watts per square meter (the speed at which the sun's energy is distributed over a given area). The energy emitted by the sun fluctuates significantly over time because of sunspots and various solar activities. This information is sourced from the SOLARIS-HEPPA project website.
The multivariate El Niño Southern Oscillation Index (MEI) indicates the intensity of the El Niño/La Niña-Southern Oscillation, a climate phenomenon in the Pacific Ocean that influences global temperatures. This information is provided by the ESRL/NOAA Physical Sciences Division. [Kaggle, 2019]

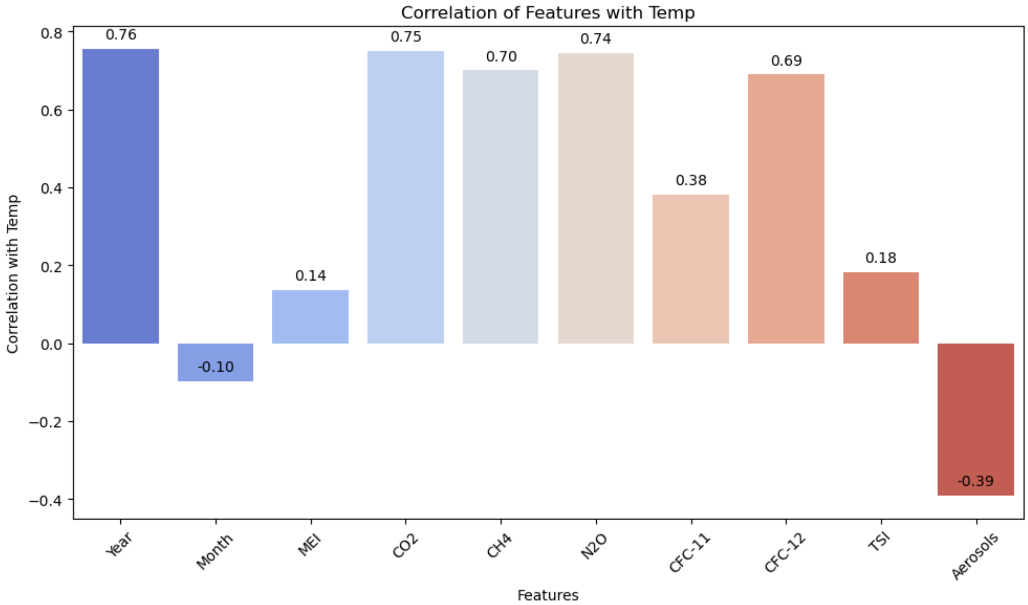

Figure 3.1: Correlation of Features with Temp-Dataset 1.

In Dataset 2, we have these variables: Year (1979-2015), Temp, CO2, CH4, N2O, CFC12, CFC11. This data comes from a combination of Berkley Earth Global Temperatures and some parts of NOAA aggi forcing datasets. The units for CO2, CH4, N2O, CFC12, and CFC11 are (Watts per square meter), indicating the contribution of each gas to radiative forcing. This unit measures the effect of each gas on the Earth's energy balance. Global Radiative Forcing (RF) refers to the change in the balance between incoming solar radiation (energy from the Sun) and outgoing infrared radiation (energy emitted by the Earth) due to factors such as greenhouse gases, aerosols, and other anthropogenic and natural influences. Radiative forcing quantifies the impact of these changes on the Earth's climate system and is a key concept in climate science used to understand how different factors influence global temperatures and climate change [Stocker et al., 2013]. Positive radiative forcing (e.g., from increased greenhouse gases like CO2) leads to warming, while negative radiative forcing (e.g., from aerosols reflecting sunlight) leads to cooling. The radiative forcing concept helps scientists measure and predict how human activities and natural processes affect the global climate [Ramaswamy et al., 2001]. The target variable (Temp) here is the Land Average Temperature regarding May in each year.

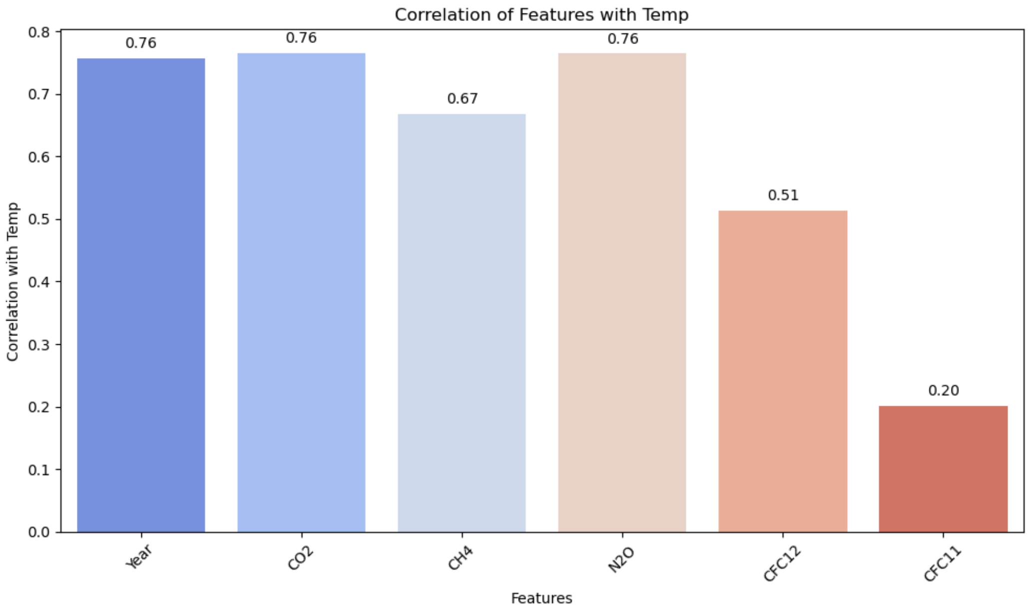

Figure 3.2: Correlation of Features with Temp-Dataset 2.

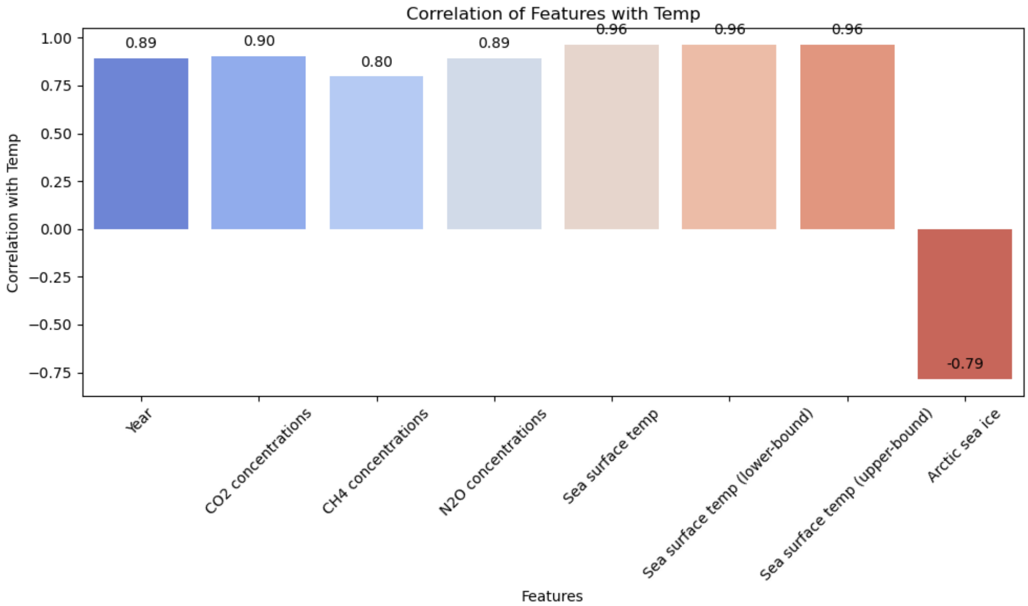

Figure 3.3: Correlation of Features with Temp-Dataset 3.

The variables in Dataset 3 are Year (1979-2015), CO2 concentrations, CH4 concentrations, N2O concentrations, Sea surface temp, Sea surface temp (lower-bound), Sea surface temp (upper-bound), Arctic sea ice, Temp. The Temp here is Earth's surface temperature change (land and ocean) in °F. The data comes from NOAA and EPA.

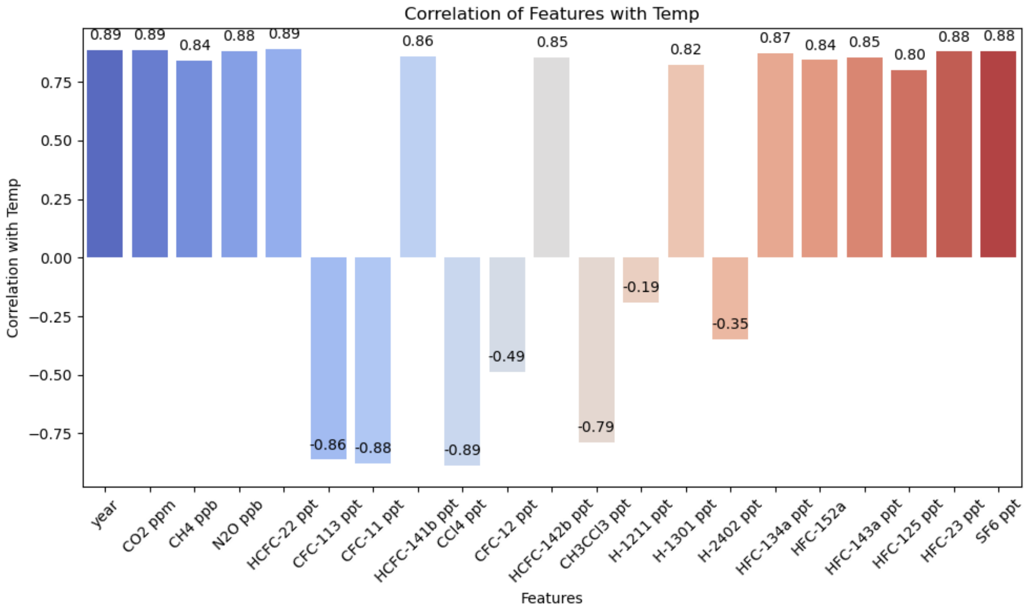

Figure 3.4: Correlation of Features with Temp-Dataset 4.

Dataset 4 have these variables: year (1992-2022), CO2 ppm, CH4 ppb, N2O ppb HCFC-22 ppt, CFC-113 ppt, CFC-11 ppt, HCFC-141b ppt, CCl4 ppt, CFC-12 ppt, HCFC-142b ppt, CH3CCl3 ppt , H-1211 ppt, H-1301 ppt, H-2402 ppt, HFC-134a ppt, HFC-152a, HFC-143a ppt, HFC-125 ppt, HFC-23 ppt, SF6 ppt, Temp. The data comes from NOAA and the Temp is 'Global Land and Ocean May Average Temperature Anomalies' in degrees Celsius.

Dataset 5 variables: Year (1980-2022), CO2, CH4, N2O, CFC, HCFCs, HFCs, Total (radiative forcing in Watts per square meter), Total CO2-eq (ppm), AGGI 1990 = 1, change (percent), Temp. The Annual Greenhouse Gas Index (AGGI) tracks the increasing amount of heat being added to the atmosphere by human-related greenhouse gas (GHG) emissions. Annual AGGI change (in percent) is calculated relative to 1990. The 'CFC*' grouping includes some other long-lived gases that are not CFCs (e.g., CCl4, CH3CCl3, and Halons), but the CFCs account for the majority (95 percent in 2022) of this radiative forcing. The 'HCFC' grouping includes the three most abundant of these chemicals (HCFC-22, HCFC-141b, and HCFC-142b). The 'HFC*' grouping includes the most abundant HFCs (HFC-134a, HFC-23, HFC-125, HFC-143a, and HFC-152a) and SF6 for completeness, although SF6 only accounted for a small fraction of the radiative forcing from this group in 2022 (13 percent). Temp is Global Land and Ocean May Average Temperature Anomalies and its unit is Degrees Celsius.

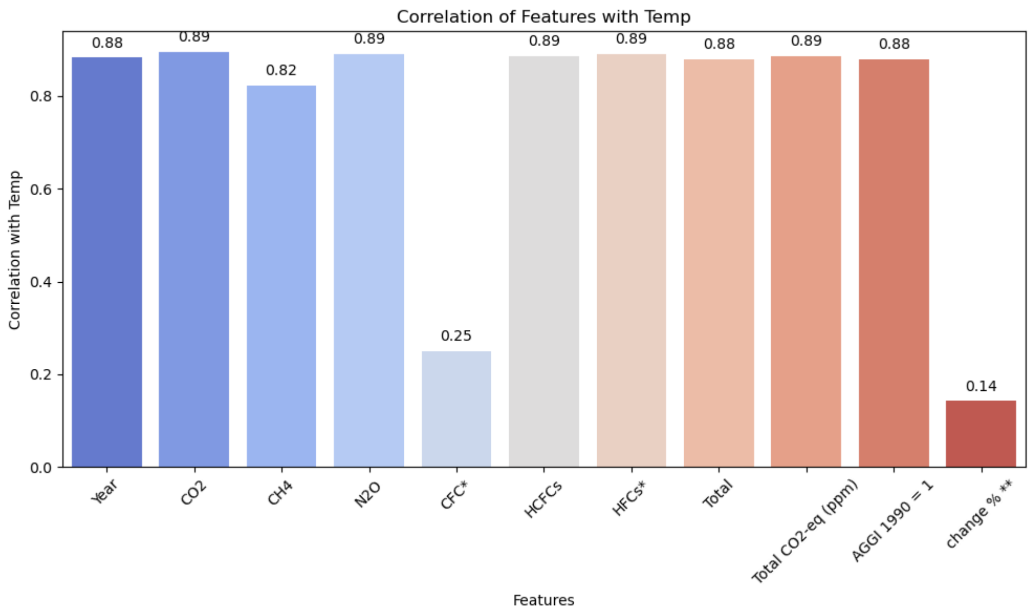

Figure 3.5: Correlation of Features with Temp-Dataset 5.

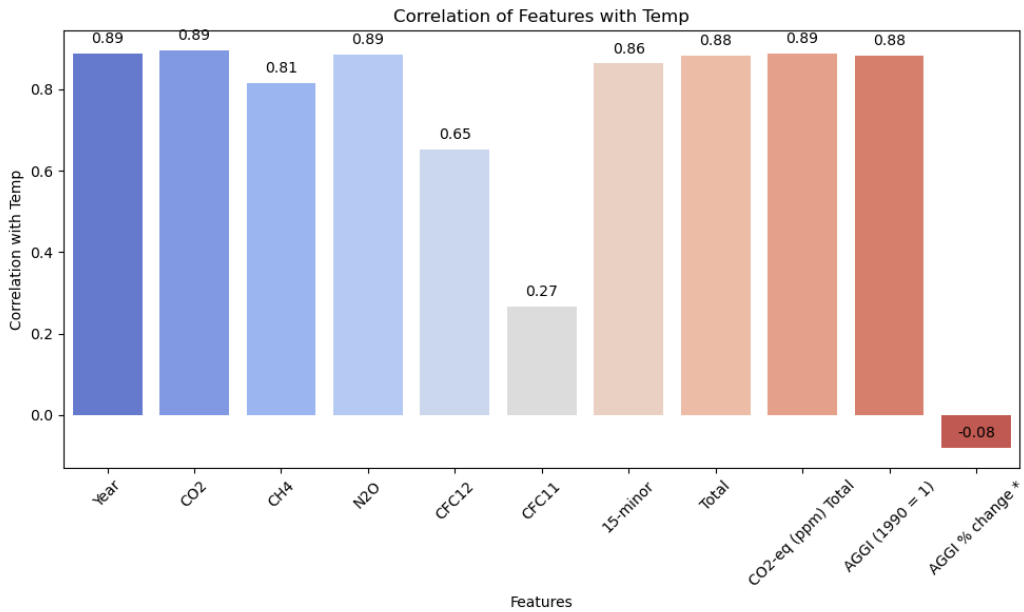

Figure 3.6: Correlation of Features with Temp-Dataset 6.

Dataset 6 variables: Year (1980-2022), CO2, CH4, N2O, CFC12, CFC11, 15-minor, Total, CO2-eq (ppm)Total, AGGI (1990 = 1), AGGI (percent change), Temp. The 'Temp' is Earth's surface temperature changes in Fahrenheit. Data source: NOAA and EPA.

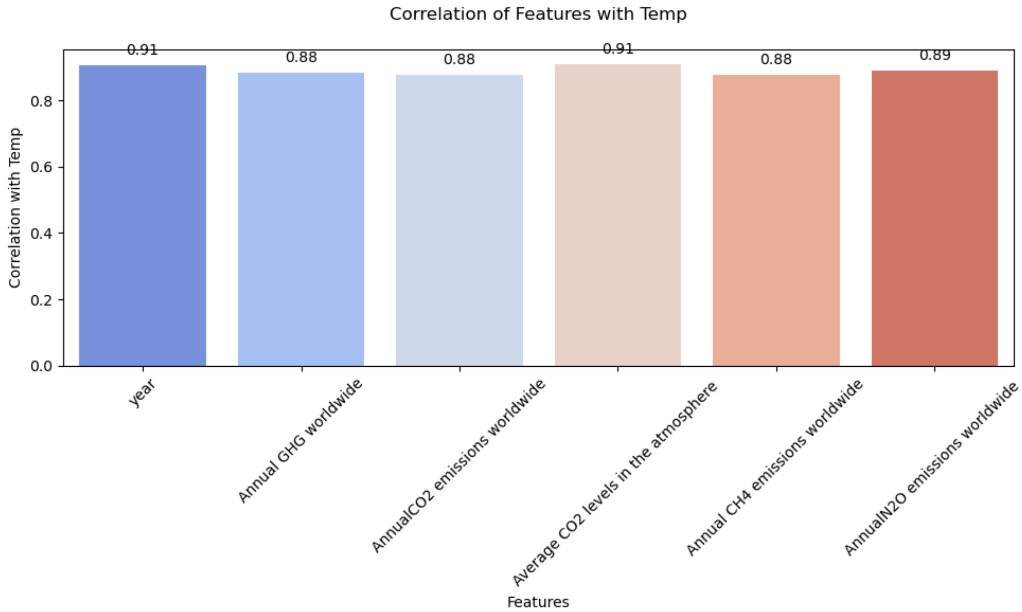

Figure 3.7: Correlation of Features with Temp-Dataset 7.

Dataset 7 has these variables: year (1990-2022), Annual greenhouse gas emissions worldwide (in billion metric tons of CO2 equivalent), Annual carbon dioxide (CO2) emissions worldwide (in billion metric tons), Average carbon dioxide (CO2) levels in the atmosphere worldwide (in parts per million), Annual methane (CH4) emissions worldwide (in million metric tons of CO2 equivalent), Annual nitrous oxide (N2O) emissions worldwide (in million metric tons of carbon dioxide equivalent), Temp. We have global average land-sea temperature anomaly relative to the 1961-1990 average temperature baseline as 'Temp' here. The information is sourced from the Met Office of the United Kingdom and encompasses both air and sea surface temperatures in the Northern and Southern Hemispheres [Ritchie et al., 2023].

Dataset 8 has these variables: Year (2001-2022), Month, Temp, co2, ch4, n2o, sf6, relative co2, relative ch4, relative n2o, relative sf6. This dataset is a combination of NOAA - Global Monitoring Laboratory's studies on greenhouse gases. The dataset focuses on the monthly tracking of the atmospheric levels and trends of key long-term climate change drivers, including carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O), as well as sulfur hexafluoride (SF_6).
relative temp : relative temperature values according to reference temperature, relative co2 : relative co2 value according to next measurement, relative ch4 : relative ch4 value according to next measurement, relative n2o : relative n2o value according to next measurement, relative sf6 : relative sf6 value according to next measurement.
CO_2 information is presented as a dry air mole fraction, which is calculated by dividing the number of carbon dioxide molecules by the total number of molecules in the air, including CO_2 itself, once water vapor has been eliminated. This mole fraction is represented in parts per million (ppm).
CH_4: Methane is presented as a "dry air mole fraction," which is defined as the ratio of methane molecules to the overall number of molecules in the sample once water vapor has been eliminated. The mole fraction is represented in nmol mol-1, commonly referred to as "ppb" (parts per billion; 1 ppb signifies that one molecule out of every billion in an air sample is CH_4).
N_2O: Nitrous oxide is expressed in terms of a "dry air mole fraction," defined as the number of nitrous oxide molecules compared to the total number of molecules in the sample after removing water vapor. This mole fraction is represented in nmol mol-1, commonly referred to as "ppb" (parts per billion; 1 ppb signifies that one molecule out of every billion in an air sample is N_2O).
SF_6: Sulfur hexafluoride is described as a "dry air mole fraction," which is

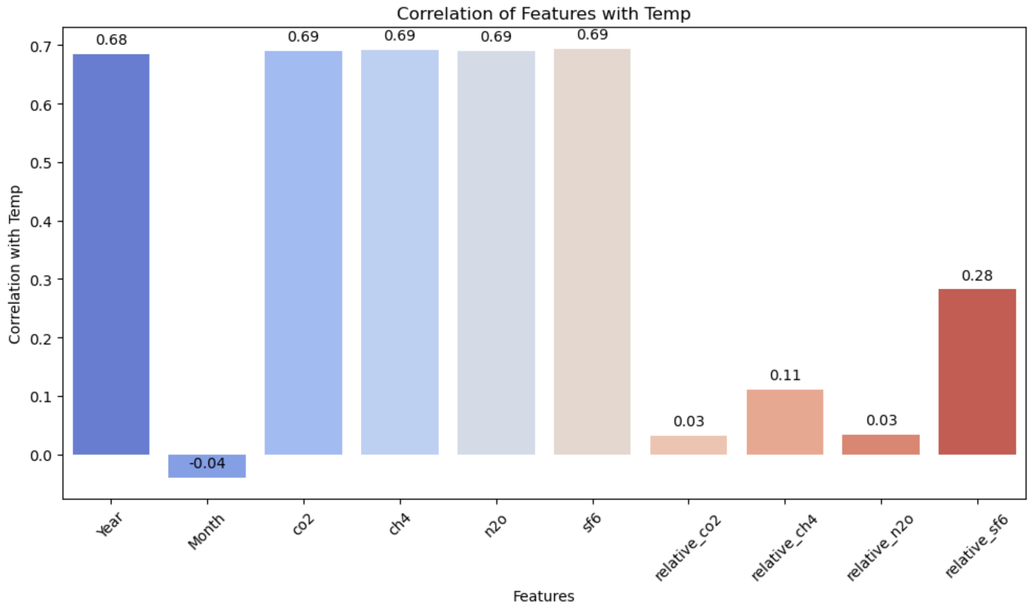

Figure 3.8: Correlation of Features with Temp-Dataset 8.

the ratio of the number of sulfur hexafluoride molecules to the total number of molecules in a sample, excluding water vapor. This mole fraction is represented in pmol mol-1, commonly referred to as "ppt" (parts per trillion; 1 ppt signifies that one molecule of SF6 exists among every trillion molecules in an air sample) [Saritas, 2022].

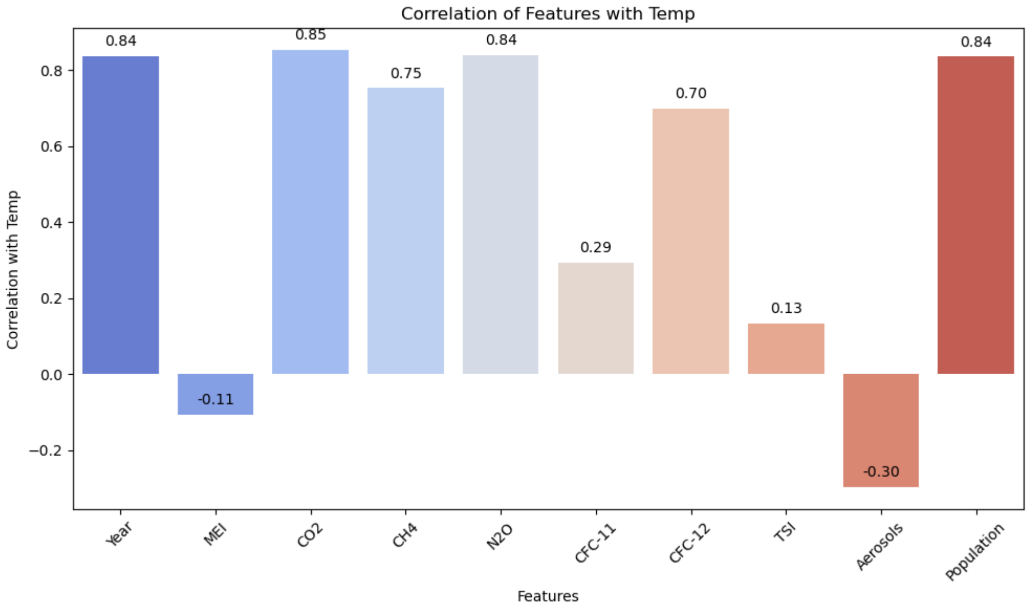

Figure 3.9: Correlation of Features with Temp-Dataset 9.

Dataset 9 variables: Year(1983-2008), MEI, CO2, CH4, N2O, CFC-11, CFC-12, TSI, Aerosols, Population, Temp. The target variable (Temp) here is the annual average temperature worldwide in Fahrenheit. Values are regarding December in each year. The data comes from Kaggle and EPA.

Dataset 10 variables: year (1990-2016), co2, coal co2, gas co2, oil co2, other industry co2, total ghg, methane, nitrous oxide, population, Temp. CO2 emissions: this information is obtained from the Global Carbon Project, which generally publishes an annual update on CO2 emissions. Greenhouse gas emissions (including methane and nitrous oxide): this information is gathered from the CAIT Climate Data Explorer and downloaded from the Climate Watch Portal. Other variables: this data is sourced from various sources, including the United Nations, World Bank, Gapminder, Maddison Project Database, and others. [Dias, 2023]. 'Temp' is Earth's surface temperature (land and ocean) [EPA, 2024].

The Year variable is a continuous measurement of time that progresses linearly, with each increment representing one uniform year. In datasets where each year is represented by 12 months, this variable effectively captures long-term trends, such as global warming. Encoding the Year variable is generally unnecessary unless it is required for non-linear modeling. As a result, in these datasets, the numeric form of the Year is adequate for representing temporal trends. The *Month* variable, in contrast, is cyclic. It repeats every 12 months, where January ($M = 1$) and December ($M = 12$) are adjacent. Treating *Month* as a simple integer variable (e.g., 1 for January to 12 for December) introduces a discontinuity at the December–January boundary, which can lead to incorrect interpretations in models. Encoding the *Month* variable to reflect its cyclic nature is essential for capturing seasonal patterns.

The most effective way to encode *Month* is using **trigonometric encoding** (Sørensen et al., 2018). This method maps months onto a unit circle, representing their periodicity while preserving the relationship between adjacent months. The encoding formulas are:

$$\text{Month\_sin} = \sin\left(\frac{2\pi \cdot M}{12}\right) \tag{3.1}$$

$$\text{Month\_cos} = \cos\left(\frac{2\pi \cdot M}{12}\right) \tag{3.2}$$

Here:

- M is the numeric representation of the month ($M \in \{1, 2, \ldots, 12\}$).
- The factor $\frac{2\pi}{12}$ maps the months onto a circle, ensuring that January ($M = 1$) and December ($M = 12$) are close to each other.

This encoding creates a two-dimensional representation of *Month* on a unit circle. The sine and cosine values allow models to capture both the direction and periodicity of the months. Importantly, the two-dimensional representation avoids the loss of information that would occur if only one trigonometric function were used [Hyndman and Athanasopoulos, 2018].

Those correlations are just pairwise correlations between each feature and temperature. However, when other features contribute and combine in the real world, their combinations influence the target variable differently.

3.2 Data Preprocessing

To obtain datasets with diverse features that impact global temperature as the target variable, we combine and merge datasets related to greenhouse

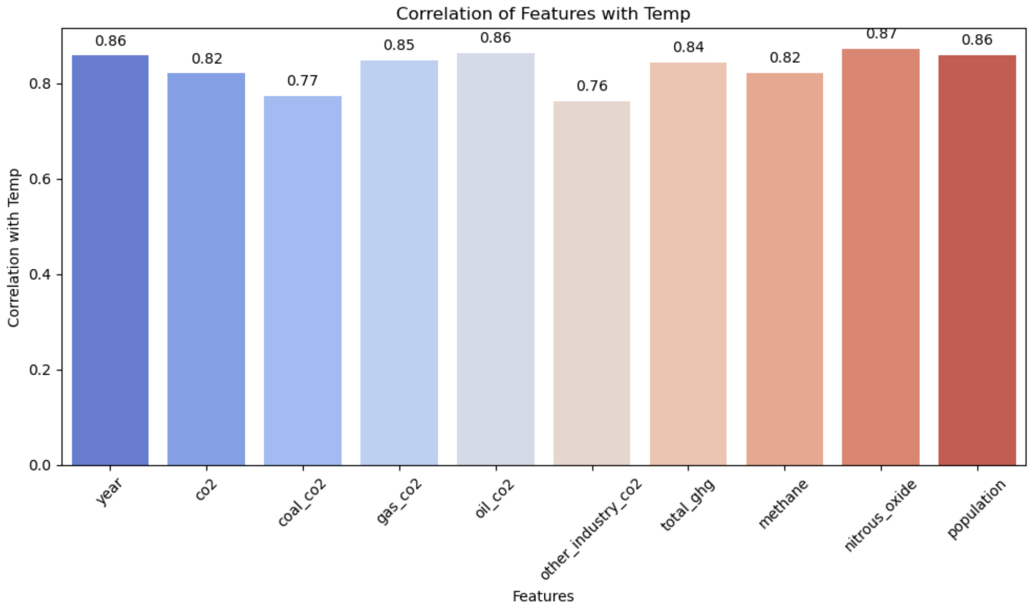

Figure 3.10: Correlation of Features with Temp-Dataset 10.

gases and global temperature anomalies with matching recorded dates. Additionally, we meticulously clean the gathered datasets to ensure that there are no missing values and to prepare them properly for the machine learning models. The models also use the following feature selection technique and training-test split method for preprocessing.

3.2.1 Feature Selection Technique

Identifying a subset of important input variables is the goal of feature selection, as these variables have the most impact on predicting the target variable. Eliminating irrelevant or redundant features enhances the model's performance, reduces overfitting, and cuts down on computation time [Guyon and Elisseeff, 2003]. A method for wrapping in feature selection is a technique that assesses feature subsets based on a predictive model's performance. In this method, the feature selection process "wraps around" the model by evaluating subsets of features through training the model and evaluating its performance (such as accuracy or error). The wrapper method entails exploring various feature combinations and choosing the subset that yields the best model performance. The process of using wrapper methods involves high computational intensity as it requires training the model multiple times on various feature subsets. Nevertheless, wrapper methods frequently yield superior outcomes compared to other feature selection techniques due to their consideration of feature interactions and their impact on the model's predictive performance [Kohavi and John, 1997].

3.2.2 Training-Test Split

The Training-Test split is an important technique for assessing the effectiveness of predictive models. This approach involves splitting the whole dataset into two separate datasets, one for training called training set and one for test called test set. The training set is utilized to train the model, while the test set is set aside for evaluating the model's performance on unfamiliar data. The division enables a fair evaluation of the model's capacity to apply to unfamiliar data [Kuhn and Johnson, 2013]. It is typical to divide the dataset, allocating 70-80 percent for training and 20-30 percent for testing. [James et al., 2013] The main goal of the training-test split is to avoid overfitting, which occurs when a model becomes too focused on the training data and doesn't perform well on new data. Keeping the test data separate helps replicate real-world situations where the model faces unfamiliar data, ensuring that the model's assessment reflects its capacity to generalize rather than just its performance on the training data [Goodfellow et al., 2016].

For the Train-Test split, first, the dataset is randomly shuffled. This randomness helps ensure that the training and test sets are representative of the overall dataset. After shuffling, the dataset is divided into two parts: 20 percent for the test set and 80 percent for the training set. Randomly splitting the data guarantees that both the train and test sets reflect the overall dataset. If the split lacks randomness, it could lead to bias.

3.3 Machine Learning Regressors

In this study, we apply several machine learning regressors to build the models, including Linear regression, Random Forest regressor, K-neighbour regressor, XGboost regressor, and Support Vector regressor.

3.3.1 Linear Regression

Linear regression is a straightforward and popular statistical technique to model the connection between a dependent variable and one or more independent variables; this model assumes that the relationship between the variables is linear, and it endeavours to determine a straight line (in simple linear regression) or a hyperplane (in multiple linear regression) that minimizes the sum of squared errors between the observed and predicted values [Montgomery et al., 2012].

3.3.2 Random Forest Regressor

Random Forest is an ensemble learning technique that creates numerous decision trees while training and producing the average prediction from the individual trees for regression tasks; the method minimizes the chance of overfitting since it combines the outcomes of multiple trees, which capture various aspects of the data's variability [Breiman, 2001].

3.3.3 K-Neighbours Regressor

The K-Neighbors Regressor constitutes a non-parametric regression algorithm categorized within the K-nearest neighbours (KNN) framework. In contrast to traditional parametric models, it does not depend on making assumptions about the distribution of the underlying data. Instead, it prioritizes the similarity between data points. Consequently, it is particularly adept at handling complex and unfamiliar data structures [Altman, 1992]

3.3.4 XGBoost Regressor

XGBoost has become extremely popular as a highly effective machine learning technique. It works by creating a powerful regressor with strong prediction capabilities using a sequence of weak tree-based regressors with low efficiency [Chen and Guestrin, 2016]. XGBoost uses a recursive binary splitting method to choose the optimal split at each stage for building the best model, and due to its tree-based structure, XGBoost is not affected by outliers and, similar to numerous boosting techniques, it is resistant to overfitting, significantly aiding in model selection [Zhang et al., 2020]

3.3.5 Support Vector Regressor

Support Vector Machines (SVM) are algorithms that utilize hyperplanes for creating regressions. In essence, the algorithm aims to distinguish different types of data by employing a hyperplane with the widest margin between the groups in a multi-dimensional space. If a data point falls outside the margin, there will be a penalty that influences the optimality of the hyperplane choice. SVM can employ various kernels, or different methods for determining the hyperplane in a high-dimensional space. Support Vector Regression (SVR) is an expansion of this concept, generating a regression based on the principles of SVM [Zheng, 2018].

3.4 Metaheuristic Optimization Algorithms for Feature Selection

Optimization algorithms based on population are among the most effective algorithms in the category of stochastic methods; these algorithms have drawn inspiration from a range of swarm intelligence phenomena, the natural behaviors of animals and insects, the principles of physics, the behaviors of players and rules in different games, and the principles of evolution [Abualigah et al., 2021b]. The search for the best solution in optimization algorithms begins by generating a set of feasible solutions based on the problem's constraints; these initial solutions are then enhanced through the algorithm's stages and an iterative process, and the optimal solution for the optimization problem is determined once the algorithm has been fully executed [Trojovský and Dehghani, 2022]. Swarm Intelligence (SI) algorithms are increasingly being used to address Feature Selection (FS) [Blum and Li, 2008]. Their nature-inspired backgrounds offer diverse approaches and strategies, making them suitable for optimization problems, including feature selection. SI has been demonstrated to be an effective method for solving complex computational problems [Garey and Johnson, 1979]. As identifying the best subset of features is definitely a challenge, In recent years, SI algorithms have become increasingly popular [Blum and Li, 2008]. We use the following natural-inspired and swarm intelligence optimization algorithms for feature selection.

3.4.1 Genetic Algorithm (GA)

GA is a natural-inspired search algorithm that takes inspiration from natural selection, and it develops a group of potential solutions to an optimization problem using operations similar to genetic processes like selection, crossover, and mutation [Holland, 1975].

3.4.2 Arithmetic Optimization Algorithm (AOA)

The Arithmetic Optimization Algorithm (AOA) is also nature-based and inspired by fundamental arithmetic operators (addition, subtraction, multiplication, and division) to balance exploration and exploitation and discover optimal solutions in a search space. [Abualigah et al., 2021a].

3.4.3 Grasshopper Optimization Algorithm (GOA)

GOA draws inspiration from grasshoppers' natural behaviour, specifically their swarming and movement behaviours, and it simulates the way grasshoppers search for food, mirroring the process of finding the best solution within a search space [Saremi et al., 2017].

3.4.4 Gravitational Search Algorithm (GSA)

The GSA algorithm is based on nature and the concept of gravity and how masses interact with each other. In this approach, agents are treated as entities, and their effectiveness is evaluated based on their masses [Rashedi et al., 2009]. The attraction force between agents is harnessed to direct the exploration towards the best solutions [Rashedi et al., 2009].

3.4.5 Harris Hawks Optimization (HHO)

The HHO algorithm takes inspiration from the collaborative hunting techniques of Harris's hawks, and it emulates the element of surprise and group coordination displayed by the hawks when capturing prey and applies this behaviour to an optimization search process [Heidari et al., 2019].

3.4.6 Grey Wolf Optimizer (GWO)

The concept of GWO comes from grey wolves' social structure and hunting tactics, and the algorithm replicates the hierarchy of wolf packs and their methods of encircling, chasing, and attacking prey to investigate and capitalize on the search space [Mirjalili et al., 2014].

3.4.7 Whale Optimization Algorithm (WOA)

WOA draws inspiration from the bubble-net hunting technique employed by humpback whales; through this algorithm, the encircling mechanism and bubble-net hunting behaviour are simulated to find the best solution within the search space [Mirjalili and Lewis, 2016].

3.4.8 Particle Swarm Optimization (PSO)

PSO is inspired by the social behaviour of birds flocking or fish schooling; it involves a group of particles (candidate solutions) that navigate the search space, guided by their individual best-known position and the best-known positions of other particles [Kennedy and Eberhart, 1995].

3.4.9 Coati Optimization Algorithm (COA)

The Coati Optimization Algorithm (COA) is a swarm intelligence technique that imitates the behavior of coatis in their natural environment. The core concept of COA is to replicate two innate behaviors exhibited by coatis: 1. their method of pursuing and capturing iguanas and 2. their escape response when faced with predators [Dehghani et al., 2023].

3.4.10 Dung Beetle Optimizer (DBO)

DBO draws inspiration from the actions of dung beetles, particularly their habit of rolling dung balls, which they use for nourishment and procreation; the algorithm translates this method into a tactic for tackling optimization issues by emphasizing both exploration and exploitation [Xue and Shen, 2022].

3.4.11 Pelican Optimization Algorithm (POA)

POA is based on pelicans' natural behaviour, and the main idea in the design of POA is to model the behaviour and strategy of pelicans during hunting, an intelligent process that has made these birds skilled hunters [Trojovský and Dehghani, 2022].

3.5 Combination

Each optimization algorithm as a feature selector will be combined with each machine learning regressor to create various models. Since we have 11 optimizers and 5 regressors, we will generate 55 different models for comparison.

3.6 Evaluation Metrics

We assess the models using various evaluation criteria, such as average, median, standard deviation, best, worst of R-squared, feature ratio and mean squared error. This will be done across 10 datasets and over 10 runs to compare the results.

3.6.1 R-squared

R-squared ($\mathbf{R^2}$) is a statistical metric that indicates the percentage of the variance in a dependent variable described by an independent variable or variables in a regression model. It is also called the coefficient of determination and gives a sense of how well the model fits. A higher $\mathbf{R^2}$ value indicates better performance [Montgomery et al., 2012]. This is the R-squared formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.3}$$

Where:

- y_i: is the actual value,
- $\hat{y}_i$: is the predicted value,
- $\bar{y}$: is the mean of the actual values,
- n: is the number of data points.

The numerator

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

represents the **residual sum of squares** (errors between actual and predicted values). The denominator

$$\sum_{i=1}^{n}(y_i - \bar{y})^2$$

is the **total sum of squares** (errors between actual values and their mean). The formula compares how well the model predicts compared to a baseline model that always predicts the mean [Draper and Smith, 1998].

Average, Median, STD, Best, Worst

We use the average of R-squared over 10 runs and the median of R-squared over 10 runs, which is a resilient measure of central tendency that reflects the middle value of a collection of R-squared values when sorted from smallest to largest [Siegel, 1982]. In contrast to the mean, the median is less affected by outliers, which makes it a more dependable gauge of typical model performance across various datasets or models, especially when dealing with skewed data [Siegel, 1982]. Also, we apply the standard deviation of R-squared that measures the variability of the R-squared values across different datasets from their average, indicating the consistency of R-squared values among the datasets [Montgomery et al., 2012, Wooldridge, 2015]. The best and worst R-squared the model achieves through the 10 runs are other metrics.

3.6.2 Feature ratio

The feature ratio in feature selection is the proportion of selected features to the total number of features present in the dataset [Guyon and Elisseeff, 2003]. The efficiency and effectiveness of a feature selection process are critically evaluated using this ratio. Selecting fewer, more relevant features is typically indicated by a lower feature ratio, making it crucial in the evaluation process [Chandrashekar and Sahin, 2014]. Maintaining a high feature ratio may suggest that numerous features are required to uphold model performance, possibly indicating redundancy or irrelevance among certain features [Guyon et al., 2002]. Reducing the number of features while maintaining performance helps improve computational efficiency and avoid overfitting. The feature ratio ensures that the model is simpler and easier to interpret while retaining the most relevant information[Guyon and Elisseeff, 2003]

3.6.3 Mean Squared Error (MSE)

The Mean Squared Error (MSE) evaluates how well a regression model performs by computing the average of the squared variances between the predicted values and the actual values. A smaller MSE suggests that the model fits the data more effectively. The formula for the Mean Squared Error is:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.4}$$

Where:

- n = number of data points
- y_i = actual value for data point i
- $\hat{y}_i$ = predicted value for data point i
- $(y_i - \hat{y}_i)^2$ = squared difference between the actual and predicted values [Montgomery et al., 2012]

3.7 Comparison

We use the Friedman test to rank and compare the performance of the models in all 10 provided datasets and in each evaluation criterion among the results. The Friedman test is a non-parametric statistical test used to analyze multiple group measures, purposing to test whether the group measures have the same variance at a specific level of significance; if the null hypothesis is rejected, it indicates that the variance values of the group measures are not equal [Ishitaki et al., 2016]. The Friedman test uses 'Rank Average' by independently ranking each row (or block) of data, with each rank being assigned based on the performance of treatments within a block. [Demšar, 2006]. The total scores for each treatment are added up within all the blocks [Hollander et al., 2013]. The Friedman test ranking clearly indicates the relative performance of each treatment, with lower ranks suggesting better performance [Iman and Davenport, 1980]. The Rank Average of the Friedman Test is a robust method for evaluating and ranking multiple algorithms across multiple datasets, making it more suitable than the Wilcoxon Signed-Rank Test. The Friedman Test is specifically designed for comparing multiple treatments (or algorithms) across multiple blocks (or datasets), making it ideal for our study where we have the performance of 11 algorithms on 10 datasets. According to [Demšar, 2006], the Friedman Test is particularly effective when comparing multiple algorithms because it provides a rank-based assessment that avoids assumptions about the normality of the data. In contrast, the Wilcoxon Signed-Rank Test is a non-parametric test for pairwise comparisons, primarily used to assess the difference between two related samples. It is more appropriate for comparing two algorithms on multiple datasets rather than a broader set of algorithms. When extended to multiple algorithms, it requires performing pairwise comparisons between all algorithm pairs, increasing the computational burden and the complexity of interpreting the results [Sheskin, 2004]. Furthermore, the Wilcoxon Signed-Rank Test does not provide a holistic ranking of all algorithms but rather a binary outcome for each pairwise comparison, which is less informative in contexts where a global ranking is needed.

In conclusion, the Rank Average of the Friedman Test is better suited for our results because it provides a holistic, interpretable ranking of all algorithms while being computationally efficient and robust to data distribution. For comprehensive statistical comparisons of multiple algorithms on multiple datasets, the Friedman Test remains the recommended choice, as supported by literature such as[Demšar, 2006] and [Derrac et al., 2011]

Chapter 4

Proposed Hybrid Algorithm

This chapter introduces the hybrid Arithmetic Optimization Algorithm-Coati Optimization Algorithm (AOA-COA), combining the global search strengths of AOA with the local search efficiencies of COA. It discusses their individual and combined advantages, emphasizing their utility in feature selection for climate predictions. The main takeaway is the development and justification of the hybrid AOA-COA for superior performance.

4.1 Arithmetic Optimization Algorithm (AOA)

The Arithmetic Optimization Algorithm (AOA) is a metaheuristic optimization technique introduced by Abualigah et al. in 2021. It draws inspiration from the distribution behavior of arithmetic operators (addition, subtraction, multiplication, and division) to navigate the search space effectively. [Abualigah et al., 2021a]

AOA operates by balancing exploration and exploitation phases during the optimization process. The transition between these phases is governed by the Math Optimizer Accelerated (MOA) function, which is defined as:

$$\text{MOA}(t) = \text{MOA}_{\text{min}} + t \times \left(\frac{\text{MOA}_{\text{max}} - \text{MOA}_{\text{min}}}{T_{\text{max}}} \right) \tag{4.1}$$

where:

- t is the current iteration number,
- T_{max} is the maximum number of iterations,
- MOA_{min} and MOA_{max} are predefined minimum and maximum values of the MOA function.

The Math Optimizer Probability (MOP) is calculated as:

$$\text{MOP} = \alpha^{\left(1 - \frac{t}{T_{\text{max}}}\right)} \tag{4.2}$$

where α is a sensitivity control parameter.

The position update equations for the exploration and exploitation phases are:

Exploration Phase:

$$X_i^{t+1} = X_i^t + r_1 \times \left(\frac{X_{\text{best}}^t - X_i^t}{|X_{\text{best}}^t - X_i^t| + \epsilon} \right) \times \text{MOP} \times (ub_j - lb_j) + lb_j \tag{4.3}$$

Exploitation Phase:

$$X_i^{t+1} = X_i^t + r_1 \times \left(\frac{X_{\text{best}}^t - X_i^t}{|X_{\text{best}}^t - X_i^t| + \epsilon} \right) \times \text{MOP} \times (ub_j - lb_j) + lb_j \tag{4.4}$$

In both equations:

- X_i^t is the position of the i-th individual at iteration t,

- X_{best}^t is the position of the best solution found so far,
- r_1 is a random number uniformly distributed in [0,1],
- ub_j and lb_j are the upper and lower bounds of the j-th dimension,
- ϵ is a small constant to prevent division by zero.[Abualigah et al., 2021a]

4.1.1 Advantages

Effective Exploration: AOA excels in global search by thoroughly exploring the solution space to avoid premature convergence. This is achieved through its arithmetic operations, which allow agents to explore diverse areas of the search space.
Dynamic Balance Between Exploration and Exploitation: The Math Optimizer Accelerated (MOA) function dynamically adjusts the algorithm's focus from exploration to exploitation as iterations progress. This helps in maintaining a good balance throughout the optimization process.
Simplicity and Flexibility: AOA is simple to implement and requires few control parameters, making it versatile for a wide range of optimization problems, including continuous and discrete spaces.
Convergence: By using mathematical operators effectively, AOA can converge to global optima with a reduced risk of getting trapped in local optima, especially in the earlier stages of the search.

Algorithm 1: Pseudocode of AOA

Input: N Number of solutions,

n Number of variables,

$T_{\max}$ Number of iterations.

Output: X^* Optimal solution position,

$f(X^*)$ Best fitness value.

1. Initialize a population of N solutions randomly within the search space
2. Calculate the fitness value of each solution
3. **while** stopping criteria are not met **do**
4. **Update** the Math Optimizer Accelerated (MOA) parameter: $\mathrm{MOA}(t) = \mathrm{MOA}_{\min} + t \times \left(\frac{\mathrm{MOA}_{\max} - \mathrm{MOA}_{\min}}{T_{\max}}\right)$
5. **Calculate** Math Optimizer Probability (MOP): $\mathrm{MOP} = \alpha^{\left(1 - \frac{t}{T_{\max}}\right)}$
6. **for** $i \leftarrow 1$ to N (all solutions in the population) **do**
7. Generate a random number r_1 in [0,1]
8. **if** $r_1 < \mathrm{MOA}$ **then**
9. // **Exploration phase**
10. **Update** position of solution according to: $X_i^{t+1} = X_i^t + r_1 \times \left(\frac{X_{\text{best}}^t - X_i^t}{|X_{\text{best}}^t - X_i^t| + \epsilon}\right) \times \mathrm{MOP} \times (ub_j - lb_j) + lb_j$
11. **else**
12. // **Exploitation phase**
13. **Update** position of solution according to: $X_i^{t+1} = X_i^t + r_1 \times \left(\frac{X_{\text{best}}^t - X_i^t}{|X_{\text{best}}^t - X_i^t| + \epsilon}\right) \times \mathrm{MOP} \times (ub_j - lb_j) + lb_j$
14. **end if**
15. **end for**
16. Evaluate the fitness of each solution
17. Update the best solution found so far X^* and best fitness $f(X^*)$

end

Figure 4.1: Pseudocode of AOA

4.1.2 Disadvantages

Sensitivity to Parameter Settings: The performance of AOA can be influenced by the choice of parameters such as alpha, MOAmin, and MOA-max.[Abualigah et al., 2021a]
Premature Convergence: Like many metaheuristic algorithms, AOA may converge prematurely to local optima, especially in complex, multimodal landscapes [Abualigah et al., 2021a].

4.2 Coati Optimization Algorithm (COA)

The Coati Optimization Algorithm (COA) is a bio-inspired metaheuristic optimization technique introduced by Dehghani et al. in 2023. It emulates the social behaviors of coatis, particularly their cooperative hunting strategies and predator avoidance tactics, to solve complex optimization problems [Dehghani et al., 2023].

COA operates by modeling two primary behaviors of coatis:

1. Hunting and Attacking Iguanas:

The coati population is divided into two equal groups. The first group climbs trees to chase iguanas down, while the second group waits on the ground to catch them. This behavior enhances exploration capabilities.

2. Escaping from Predators:

When threatened, coatis flee to the nearest safe location. This behavior enhances exploitation capabilities. The mathematical representation of these behaviors is as follows:

Initialization:
A population of N coatis is initialized randomly within the search space:

$$X_i = X_{\min} + \text{rand}() \times (X_{\max} - X_{\min}), \quad i = 1, 2, \ldots, N \tag{4.5}$$

where $X_{\min}$ and $X_{\max}$ are the lower and upper bounds of the search space, respectively.

Hunting and Attacking Phase:
For each coati i, its position is updated based on the positions of the best and worst solutions:

$$X_i(t+1) = X_i(t) + r_1 \times (X_{\text{best}}(t) - X_i(t)) - r_2 \times (X_{\text{worst}}(t) - X_i(t)) \tag{4.6}$$

where r_1 and r_2 are random numbers uniformly distributed in [0,1], X_{best} is the position of the best solution, and X_{worst} is the position of the worst solution.

Escaping from Predators Phase:

Coatis move away from predators by updating their positions as follows:

$$X_i(t+1) = X_i(t) + r_3 \times (X_i(t) - X_{\text{predator}}(t)) \tag{4.7}$$

where r_3 is a random number uniformly distributed in [0,1], and X_{predator}

represents the position of the predator. [Dehghani et al., 2023]

Algorithm 2: Pseudocode of COA

Input: N Number of coatis,

n Number of variables,

t Number of iterations.

Output: X^* Optimal coati position,

$f(X^*)$ Best fitness value.

1. Initialize a population of N coatis randomly within the search space
2. Calculate the fitness value of each coati
3. **while** stopping criteria are not met **do**
4. **for** $i \leftarrow 1$ to N (all N coatis in the population) **do**
5. Divide the population into two groups (hunting and escaping)
6. **Generate random numbers** $r_1, r_2, r_3 \in [0, 1]$
7. **if** coati belongs to the hunting group **then**
8. **Update** position of coati according to:
$X_i(t+1) = X_i(t) + r_1 \times (X_{\text{best}}(t) - X_i(t)) - r_2 \times (X_{\text{worst}}(t) - X_i(t))$
9. **else** (coati belongs to the escaping group)
10. **Update** position of coati according to:
$X_i(t+1) = X_i(t) + r_3 \times (X_i(t) - X_{\text{predator}}(t))$
11. **end if**
12. **end for**
13. Evaluate the positions of individual coatis
14. Set the best fitness $f(X^*)$ value to the best solution X^*

end

Figure 4.2: Pseudocode of COA

4.2.1 Advantages

Intensive Exploitation: COA is effective in refining solutions locally. COA efficiently fine-tunes the solution to find the local optimum by mimicking the coati's hunting and escaping behaviors.
Adaptive Local Search: COA adapts to the local landscape by intensively exploring around the current best solution. This helps in exploiting the local search space more thoroughly to improve solution accuracy.
Handling Non-linearity and Complex Landscapes: The combination of COA's local search capabilities with its adaptive mechanisms makes it effective in handling complex, nonlinear optimization problems, where fine-tuning is essential.

4.2.2 Disadvantages

Premature Convergence: COA may converge prematurely to local optima, especially in complex, multimodal landscapes.
Sensitivity to Initial Conditions: The performance of COA can be influenced by the initial population distribution.
Computational Cost: For high-dimensional problems, the algorithm may require a significant number of function evaluations, leading to increased computational time.[Dehghani et al., 2023]

4.3 Proposed Hybrid AOA-COA

Hybridizing one metaheuristic with another is a widely used approach to improve the effectiveness of both algorithms. This process enables the merging of the strengths of both algorithms to achieve superior performance [Tawhid and Dsouza, 2018]. Hybridization often combines a global search algorithm with a local search one to achieve a balanced exploration-exploitation strategy. For example, hybridizing a particle swarm optimization (PSO) algorithm with a local search algorithm can enhance the precision of PSO's convergence by allowing fine-tuning of solutions. By first using a global search algorithm to explore a promising region and then refining the solution with a local search, the hybrid algorithm can achieve high-quality results more efficiently than each component algorithm could individually [Talbi, 2009].

We integrate two natural-inspired algorithms: Arithmetic Optimization Algorithm (AOA), which is the best one in our initial experiments and comparisons among the others, and the Coati Optimization Algorithm (COA), which is a swarm intelligence algorithm recently introduced in 2023, through hybridization to create a more accurate and robust algorithm for feature selection. The hybridization is entirely based on algorithms, merging characteristics of AOA and COA.

In nature-inspired optimization algorithms, exploration refers to the ability to broadly search across the solution space to identify promising areas, while exploitation focuses on refining solutions within these areas to approach an optimal result [Yang and Deb, 2009].

The research by Abualigah (2021) highlights the global solid search capabilities of the Arithmetic Optimization Algorithm (AOA) in various benchmark optimization problems and real-world applications. AOA outperformed metaheuristics like Particle Swarm Optimization, Genetic Algorithm, and Differential Evolution in locating global optima across diverse test functions. The algorithm is designed for exploration, utilizing arithmetic-based update rules that enable significant positional changes for agents, helping them avoid local optima. However, this focus on exploration can hinder its ability to refine solutions near the optimum due to a lack of precise local adjustment mechanisms [Abualigah et al., 2021a].

Additionally, the Coati Optimization Algorithm (COA) is effective for exploitation, utilizing leader-following and hunting behaviors to refine solutions through a balance of exploration and exploitation. Its strong local search capabilities enable COA to fine-tune promising solutions by making small, incremental changes, allowing it to converge on optimal or near-optimal solutions.[Dehghani et al., 2023].

It was anticipated that their combined performance will surpass that of separate algorithms. Consequently, we hybridized the global search capabilities of the Arithmetic Optimization Algorithm (AOA) with the local search functionalities of the Coati Optimization Algorithm (COA) to enhance the feature selection process.

The Arithmetic Optimization Algorithm (AOA) serves as a global search algorithm for exploring the feature space. It commences with a randomly generated binary vector, with each element indicating whether a feature is included (1) or excluded (0). The algorithm continuously updates the solution using arithmetic operations, modifying feature selection to minimize the Mean Squared Error (MSE). The Coati Optimization Algorithm (COA) improves candidate solutions provided by AOA by exploring nearby feature sets and making local adjustments. COA enhances feature selection through random movements and perturbations, refining solution positions. It aims to exploit potential areas identified by AOA to improve solutions and ensure the final feature subset is locally optimal.[Abualigah et al., 2021a, Dehghani et al., 2023]

After selecting a subset of features through the hybrid AOA-COA process, the model is trained using only the selected features and evaluated on test sets. This process repeats for 10 runs, and various performance metrics, including R-squared values (Mean, Median, STD, Best, Worst), the feature ratio and MSE are computed.

High-dimensionality and overfitting are two significant challenges encountered in predictive modeling, particularly when dealing with complex datasets such as those used in climate-related studies. High-dimensional datasets often contain a large number of features, many of which may be irrelevant or redundant. This not only increases computational complexity but also reduces model interpretability and performance. To address this, the hybrid AOA-COA algorithm are employed for feature selection to identify the most relevant subset of features, thereby reducing dimensionality while retaining critical information. The combination of AOA's global search capabilities and COA's local refinement ensures a comprehensive exploration of the feature space and precise selection of relevant features.

Overfitting poses another critical challenge, where the model captures noise and specific patterns in the training data that do not generalize to new data. This results in high training accuracy but poor performance on test datasets. To mitigate overfitting, the hybrid AOA-COA algorithm utilizes a balanced approach to exploration and exploitation. While AOA explores a wide feature space to avoid premature convergence, COA fine-tunes the selected features for enhanced precision. The use of evaluation metrics like Mean Squared Error (MSE) and R-squared across multiple datasets and runs further ensures that the model generalizes well to unseen data. Cross-validation techniques are also applied to enhance robustness and minimize model variance.

This hybrid AOA-COA approach improves model performance and accuracy by effectively exploring the search space and refining solutions. By applying AOA for broad exploration and COA for focused exploitation, the combined method strikes a balance that increases the chances of identifying high-quality solutions.

Algorithm 3: Pseudocode of Hybrid AOA-COA

Input:

N_{AOA}: Number of agents in AOA
N_{COA}: Number of coatis in COA
T_{AOA}: Number of iterations for AOA
T_{COA}: Number of iterations for COA
n: Number of dimensions (features)

Output:

X^*: Best solution found
$f(X^*)$: Best fitness value

AOA Global Search:
|
| 1. Run AOA with N_{AOA} agents and T_{AOA} iterations
| 2. Obtain best solution X_{AOA} and $f(X_{\mathrm{AOA}})$

Initialize COA:
|
| 1. Initialize N_{COA} coatis around X_{AOA}
| 2. Set $X_{\mathrm{leader}} = X_{\mathrm{AOA}}$

COA Local Search:
|
| 1. Run COA with T_{COA} iterations
| 2. Obtain refined best solution X_{COA} and $f(X_{\mathrm{COA}})$

Final Evaluation:
|
| 1. Evaluate fitness $f(X_{\mathrm{COA}})$
| 2. Return the best solution $X^* = X_{\mathrm{COA}}$ and fitness $f(X^*)$

└─ **End**

4.3.1 Combined Advantages of The Proposed Hybrid AOA-COA

Enhanced Global-Local Search Capability: The hybrid method utilizes AOA's robust global search capabilities to extensively investigate the search domain, while COA's strengths in local search are employed to fine-tune the optimal solutions. This combination improves overall optimization performance.

Reduced Premature Convergence: By switching from global exploration (AOA) to local exploitation (COA), the hybrid approach mitigates the risk of premature convergence to suboptimal solutions.

Improved Solution Accuracy: The refined search process from COA improves the accuracy of the final solution, ensuring better optimization results.

Versatility in Diverse Applications: This hybrid methodology is versatile and can be applied to various domains such as feature selection, engineering design, image processing, and more, effectively addressing both global and local optimization challenges. [Abualigah et al., 2021a, Dehghani et al., 2023].

Chapter 5

Discussion and Results

5.1 Climate Science Models and Machine Learning

Predicting global average temperature changes is crucial for understanding climate dynamics and guiding policy decisions. Among the most influential frameworks for this purpose is the cumulative carbon model, as proposed by [Matthews et al., 2009] and expanded upon by [Matthews et al., 2012]. This model highlights a near-linear relationship between cumulative CO2 emissions and global temperature anomalies, expressed as $\Delta T = \text{CCR} \cdot E$ with ΔT representing the temperature anomaly, E the cumulative carbon emissions, and CCR (Cumulative Carbon Response) is a constant representing the temperature increase per unit of cumulative CO2 emissions. This relationship has been extensively validated across scenarios and models, demonstrating its reliability for long-term climate predictions and policymaking.

Climate science models, such as Matthews' cumulative carbon framework, are particularly well-suited for predicting global temperature trends over the long term and for guiding global climate goals. These models are built upon well-established physical principles, emphasizing the near-linear relationship between cumulative carbon dioxide emissions and global temperature anomalies. Their simplicity and transparency make them highly effective for capturing long-term warming trends and for establishing emissions targets aligned with stabilization thresholds, such as limiting warming to 1.5°C or 2°C [Matthews et al., 2009, 2012]. By focusing on the cumulative nature of emissions, these models provide robust predictions for long-term trajectories and are widely adopted in international climate agreements and policy frameworks. However, their abstraction of complex processes means they are less adept at accounting for localized feedbacks, short-term variability, and non-CO2 forcings, such as aerosols and methane. As a result, their utility is generally more aligned with broad, global-scale analysis rather than localized or near-term dynamics.

In contrast, machine learning models excel in analyzing complex datasets and capturing non-linear relationships among diverse climate variables. These capabilities make machine learning particularly advantageous for short-term trends and regional predictions. By incorporating real-time data and accounting for rapid changes in variables such as aerosols, solar radiation, and ocean currents, machine learning models provide adaptive and precise forecasting for shorter timescales. Furthermore, machine learning's ability to work with high-resolution data allows it to capture the unique characteristics of specific regions, making it invaluable for generating localized climate projections. For example, these models can address phenomena like urban heat islands or regional atmospheric dynamics, which are critical for developing regional adaptation strategies. The insights provided by machine learning models enable policymakers to identify vulnerable areas and implement targeted adaptation measures, such as optimizing water resource management or enhancing resilience to heat stress [Reichstein et al., 2019, Rolnick et al., 2019]. However, machine learning models are not without limitations. They often lack the interpretability and theoretical grounding of physics-based climate science models. Their reliance on high-quality data and potential for overfitting also constrain their applicability for long-term global trends.

The following table presents the performance evaluation metrics for three models on single datasets for global temperature anomaly prediction: Linear Regression (LR), Support Vector Regressor (SVR) which are basic machine learning models, and Climate Science Model (CSM) based on cumulative carbon response. The metrics are R-squared, Mean Squared Error (MSE), and Root Mean Squared Error (Root MSE).

From the table, the R-squared value is highest for the Climate Science Model (0.94), indicating that it explains a greater proportion of the variance in the data compared to the other models. This suggests that the Climate Science Model is better suited for capturing long-term global temperature trends, as it likely integrates domain-specific knowledge and physical laws relevant to climate dynamics. On the other hand, machine learning models like Linear Regression and Support Vector Regressor show slightly lower R-squared values (0.86 and 0.87, respectively), implying that while they are effective, they may be better suited for capturing localized or short-term variations rather than comprehensive long-term trends.

In terms of error metrics, MSE and Root MSE values are lowest for the Support Vector Regressor (0.0085 and 0.0922), followed by Linear Regression

(0.009 and 0.095). These lower error values indicate that machine learning models perform slightly better in minimizing prediction errors, particularly in the short term or for smaller-scale applications. The Climate Science Model has a higher MSE (0.0170) and Root MSE (0.1306), which could be due to its focus on broader, long-term predictions rather than precise localized accuracy. However, it is important to note that these findings are based on a single dataset, and a thorough comparison should involve evaluating multiple datasets and a variety of machine learning and climate science models. This comparison emphasizes the complementary strengths of these models. In summary, while climate science models are adept at long-term climate projections due to their basis in physical laws, machine learning models offer advantages in short-term and localized forecasting through pattern recognition in historical data. The development of hybrid models seeks to unify these strengths, potentially enhancing predictive capabilities across various temporal and spatial scales [Randall et al., 2007, Reichstein et al., 2019].

Evaluation Metrics / Models	LR	SVR	CSM
R-squared	0.86	0.87	0.94
MSE	0.0090	0.0085	0.0170
Root MSE	0.0950	0.0922	0.1306

Table 5.1: Evaluation Metrics for LR, SVR, and CSM Models

The complementary strengths of these approaches highlight their distinct yet interrelated roles in climate science. Climate science models, like Matthews', provide a robust and theoretically grounded framework for understanding long-term global warming and setting emissions targets. Machine learning, on the other hand, adds granularity and adaptability, making it particularly useful for addressing localized and short-term climate variability. Together, these methods form a comprehensive toolkit, enabling robust predictions of long-term climate trends while addressing the complexities of regional adaptation strategies. By integrating these approaches, researchers can leverage the global insights of scientific models while enhancing precision and relevance with machine learning, ultimately improving both climate science and policy-making [Matthews et al., 2009, 2012, Reichstein et al., 2019, Rolnick et al., 2019].

Predicting global average temperature requires models capable of handling the complexities of climate systems and large datasets. While climate science models such as the cumulative carbon framework offer clear insights into long-term global warming trends, they are limited in their ability to incorporate the intricate, non-linear relationships present in climate systems. Machine learning (ML) models complement these frameworks by leveraging large datasets and capturing complex dynamics. However, ML models face challenges with interpretability and computational efficiency, particularly when handling high-dimensional data. This study addresses these challenges by integrating nature-inspired optimization algorithms into a hybrid approach for feature selection, improving predictive performance while retaining interpretability.

The proposed hybrid feature selection method combines the Arithmetic Optimization Algorithm (AOA) for global search with the Coati Optimization Algorithm (COA) for local refinement. AOA, as a global search algorithm, effectively explores the feature space, generating an initial subset of features based on their contribution to minimizing the model's mean squared error (MSE). However, AOA's exploratory nature can sometimes overlook finer adjustments near optimal solutions. To address this, COA refines these initial selections through local search, enhancing precision and ensuring the final subset of features is locally optimal. This synergy between AOA's exploration capabilities and COA's exploitation mechanisms balances global and local optimization, addressing the trade-off between generality and specificity in feature selection.

High-dimensional datasets pose significant challenges in climate-related studies, often containing numerous irrelevant or redundant features that increase computational complexity and risk overfitting. By reducing dimensionality, the hybrid AOA-COA approach minimizes these risks while preserving critical information. For instance, the approach evaluates features based on their predictive importance using fitness functions that prioritize MSE minimization and R-squared improvement. By repeatedly optimizing feature subsets over multiple runs, the method ensures robustness and reduces model variance.

In this study, the hybrid AOA-COA algorithm was tested across ten datasets using various regressors, including Linear Regression, Random Forest, K-Neighbors, XGBoost, and Support Vector Regressors. The results demonstrated consistent improvements in R-squared, feature ratio, and MSE across all datasets. The hybrid algorithm outperformed other metaheuristics that participated in this study, highlighting its reliability and adaptability. For example, when used with XGBoost and Random Forest regressors, the hybrid approach achieved the highest predictive accuracy while maintaining a reduced feature ratio, underscoring its efficiency in feature selection optimization.

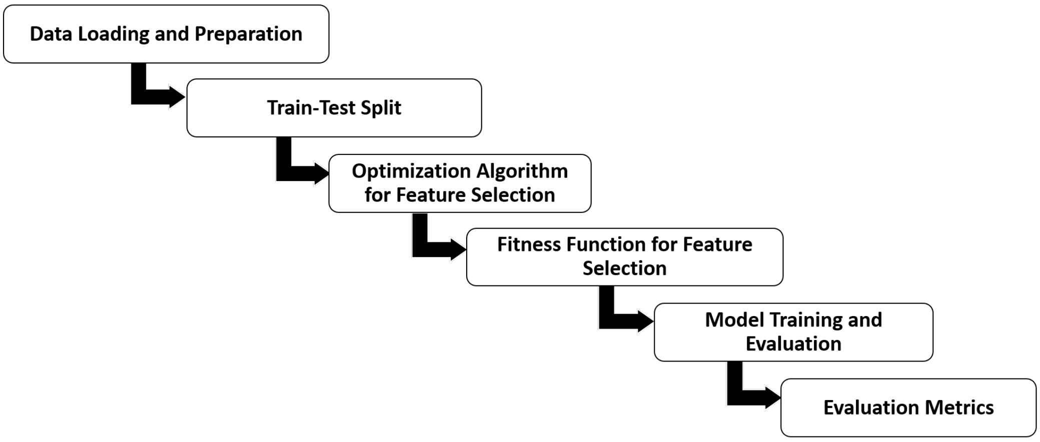

Figure 5.1: Implementation Process

5.2 Implementation Process

The following picture shows the implemented process; after that, we explain each step.

Step 1: Data Loading and Preparation

The first step is loading the dataset and splitting it into features (input variables) and the target variable "Temp." The dataset is loaded using the Panda's library, and the target variable is separated from the rest of the data, which means that we tell the machine which variable is the target to predict. This step is crucial because the dataset needs to be divided into features (inputs) and the target (output) to train a predictive model.

Step 2: Train-Test Split

The data is split into training and test sets using the train-test-split function from Scikit-learn library. This ensures that the model is evaluated on unseen data to simulate how it would perform in the real world. Typically, 80 percent of the data is used for training, and 20 percent is used for testing [Hashemitaheri et al., 2020, Liu et al., 2023]:

So, we have a "Training set" that is used to train the model and a "Test set" which is used to evaluate the model's performance on unseen data; to have a same condition in all models we used 80 percent of each dataset for train set and 20 percent for test set.

The train-test split is essential in machine learning because it prevents the model from memorizing the training data and ensures it generalizes well to new data [Goodfellow et al., 2016].

Step 3: Optimization Algorithm for Feature Selection

The model applies a wrapper-based feature selection technique using the optimization algorithm where the model's performance guides the feature selection process. The algorithm aims to identify the best feature subset by iterating through multiple populations. It explores various feature combinations and aims to minimize the MSE by selecting the best subset. The process of exploring new feature subsets by exploring the complete solution space is known as "global search" or "exploration phase", and when the algorithm exploits the best solutions found so far to refine the search, it is in the "local search" or "exploitation phase." The primary focus is on exploration, as the algorithm aims to encompass a wide range of potential solutions to prevent becoming trapped in a local optimum. [Yang, 2010].

Local search involves improving a solution by examining its surroundings, such as making minor modifications to the current solution and verifying if these adjustments result in a better solution. Local search emphasizes exploitation, which means it concentrates on enhancing solutions that are already deemed satisfactory by exploring neighboring configurations.[Aarts and Lenstra, 1997]. Once an area with potential in the solution space is identified during a global search, it is possible to employ local search to enhance the solution and obtain an optimal or nearly optimal outcome.[Hoos and Stützle, 2005]. This optimization process helps find the best subset of features that minimizes the MSE of the predictive model. Wrapper methods, such as this one, are highly efficient because they assess the value of every feature subset by directly measuring model performance [Guyon and Elisseeff, 2003].

Step 4: Fitness Function for Feature Selection

The fitness function evaluates the subsets of features based on how well they help the regressor predict the target variable. The optimization algorithm uses the fitness function to determine the quality of a feature subset.

It trains the model on the selected features and calculates the mean squared error (MSE) on the test set. This approach guarantees that the model is built using only the most appropriate features, enhancing efficiency and accuracy [Kohavi and John, 1997]. The function uses cross-validation to train the model with the chosen features and calculate the Mean Squared Error (MSE). Employing cross-validation aids in obtaining a dependable assessment of model performance by dividing the data into several subsets for training and validation [Kohavi, 1995].

Step 5: Model Training and Evaluation

For each selected subset of features, the regressor is trained and evaluated using the test set. The mean squared error (MSE) and R-squared (R^2) values are calculated for test sets to assess model performance.

Step 6: Evaluation Metrics

Optimization algorithm for the feature selection process is applied 10 times to ensure robustness. After each run, the model calculates and stores the MSE, R-squared, and feature ratio. Then, it calculates summary statistics, including the mean, median, best, worst, and standard deviation of R-squared and average of MSE and feature ratio over 10 runs. This enables the feature selection method to be assessed through multiple iterations, guaranteeing that the outcomes are stable and trustworthy [Guyon and Elisseeff, 2003].

5.3 Initial Results

We used 10 datasets to ensure a comprehensive evaluation. We created 55 models by combining each feature selection algorithm with each regressor and ran the models 10 times for each dataset. We assessed the models using various evaluation criteria, such as average, median, standard deviation, best, worst of R-squared, and feature ratio. This assessment was done across 10 datasets and over 10 runs to compare the results.

We applied the Rank Average function from the Friedman test in each regressor through the 10 datasets to have the overall ranking of the models' performance in each regressor. Tables 5.2, 5.3, 5.4, 5.5, and 5.6 show the ranking of the models over all 6 evaluation criteria in each regressor.

In this stage, we aimed to identify the overlaps among the 5 machine learning regressors. The results indicated that the Arithmetic Optimization Algorithm (AOA) was the best natural-inspired algorithm, followed by Particle Swarm Optimization (PSO) and Gravitational Search Algorithm (GSA). These three algorithms ranked the highest across all evaluation metrics and all regressors compared to the others we examined. Additionally, the results showed that Harris Hawks Optimization (HHO) is the worst-performing algorithm.

Finally, we found the overall ranking over all 5 regressors and all 6 evaluation criteria, as shown in the table 5.7. The natural-inspired optimization algorithms' names are shown with the year they were introduced.

Linear Regression	**Ranking**
(AOA) 2021	1
(PSO) 1995	2
(GOA) 2017	3
(POA) 2022	4
(GSA) 2009	5
(COA) 2023	6
(DBO) 2022	7
(WOA) 2016	8
(GA) 1975	9
(GWO) 2014	10
(HHO) 2019	11

Table 5.2: Ranking Over All Evaluation Metrics with Linear Regression

Random Forest Regressor	**Ranking**
(PSO) 1995	1
(AOA) 2021	2
(GSA) 2009	3
(GOA) 2017	4
(POA) 2022	5
(COA) 2023	6
(DBO) 2022	7
(GA) 1975	8
(WOA) 2016	9
(GWO) 2014	10
(HHO) 2019	11

Table 5.3: Ranking Over All Evaluation Metrics with Random Forest Regressor

K-Neighbors Regressor	**Ranking**
(AOA) 2021	1
(PSO) 1995	2
(GSA) 2009	3
(DBO) 2022	4
(POA) 2022	5
(COA) 2023	6
(GOA) 2017	7
(GA) 1975	8
(GWO) 2014	9
(WOA) 2016	10
(HHO) 2019	11

Table 5.4: Ranking Over All Evaluation Metrics with K-Neighbors Regressor

XGboost Regressor	**Ranking**
(AOA) 2021	1
(PSO) 1995	2
(GSA) 2009	3
(GOA) 2017	4
(DBO) 2022	5
(POA) 2022	6
(GA) 1975	7
(COA) 2023	8
(WOA) 2016	9
(GWO) 2014	10
(HHO) 2019	11

Table 5.5: Ranking Over All Evaluation Metrics with XGboost Regressor

Support Vector Regressor	**Ranking**
(GSA) 2009	1
(PSO) 1995	2
(GWO) 2014	3
(AOA) 2021	4
(DBO) 2022	5
(GOA) 2017	6
(POA) 2022	7
(COA) 2023	8
(GA) 1975	9
(HHO) 2019	10
(WOA) 2016	11

Table 5.6: Ranking Over All Evaluation Metrics with Support Vector Regressor

Over All Regressors	**Ranking**
(AOA) 2021	1
(PSO) 1995	2
(GSA) 2009	3
(GOA) 2017	4
(POA) 2022	5
(DBO) 2022	6
(COA) 2023	7
(GA) 1975	8
(GWO) 2014	9
(WOA) 2016	10
(HHO) 2019	11

Table 5.7: Ranking Over All Evaluation Metrics and All Regressors

5.4 Proposed Hybrid AOA-COA's Results

We compared the results of the new hybrid AOA-COA algorithm with the top 3 algorithms (AOA, PSO, GSA) among the 11 natural-inspired algorithms and the COA that we have in the hybrid algorithm. Again, we examined it through the 10 datasets and implemented 10 runs to have the same situation and comparable results. The idea was that if the hybrid algorithm had outperformed the AOA, PSO, GSA, and COA, it would considered the best among all the algorithms presented and used. The hybrid AOA-COA showed better results over all evaluation metrics and all regressors. In addition, the results indicate that the hybrid AOA-COA surprisingly outperformed the others in terms of accuracy when considering only a combination of algorithms with the Support Vector regressor.

The following tables show all 6 evaluation criteria results over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

Average R-squared over 10 runs

Datasets	Linear Regression AOA	Linear Regression GSA	Linear Regression PSO	Linear Regression COA	Linear Regression Hybrid AOA-COA
Dataset 1	0.710563296	0.705707823	0.710517272	0.69873803	0.71088899
Dataset 2	0.641114674	0.641114674	0.637889244	0.633775945	0.641114674
Dataset 3	0.973687238	0.692969803	0.972471513	0.965917153	0.973687238
Dataset 4	0.960734784	0.910913797	0.934499945	0.91137995	0.949708423
Dataset 5	0.906302292	0.887669597	0.893786295	0.886965241	0.906302292
Dataset 6	0.843251856	0.823155804	0.829789881	0.820871144	0.836861027
Dataset 7	0.567920354	0.56162371	0.562820048	0.556021305	0.567920354
Dataset 8	0.446860296	0.437925765	0.441951248	0.434141005	0.446860296
Dataset 9	0.433885677	0.490838817	0.530841117	0.603430264	0.807697797
Dataset 10	0.641365924	0.626836569	0.66369143	0.666886815	0.66405052

Datasets	RandomForest Regressor AOA	RandomForest Regressor GSA	RandomForest Regressor PSO	RandomForest Regressor COA	RandomForest Regressor Hybrid AOA-COA
Dataset 1	0.858748098	0.851067656	0.854230503	0.844597571	0.859280676
Dataset 2	0.499596302	0.498055314	0.498055314	0.478600436	0.499112706
Dataset 3	0.944676043	0.627967342	0.942291746	0.942620913	0.942646454
Dataset 4	0.89897786	0.915614088	0.925626132	0.894431036	0.931587049
Dataset 5	0.803123531	0.812424049	0.814910272	0.796901639	0.818196925
Dataset 6	0.751411246	0.736656649	0.742503069	0.73626039	0.750407008
Dataset 7	0.585728456	0.581938	0.582653928	0.599555495	0.586030471
Dataset 8	0.791798938	0.788658948	0.789830669	0.771034151	0.795602391
Dataset 9	0.337780767	0.380342664	0.338852428	0.381750952	0.531020417
Dataset 10	0.637943787	0.66041417	0.709253641	0.684988649	0.671570658

Datasets	Kneighbors Regressor AOA	Kneighbors Regressor GSA	Kneighbors Regressor PSO	Kneighbors Regressor COA	Kneighbors Regressor Hybrid AOA-COA
Dataset 1	0.827137165	0.812486138	0.827273366	0.809426421	0.827273366
Dataset 2	0.632364832	0.632364832	0.632364832	0.608564772	0.632364832
Dataset 3	0.927844972	0.653360149	0.914127656	0.844072523	0.927844972
Dataset 4	0.885180422	0.882823417	0.875558541	0.823493282	0.885399232
Dataset 5	0.823610129	0.821256672	0.807352974	0.785038907	0.823610129
Dataset 6	0.672339753	0.672339753	0.672339753	0.672339753	0.672339753
Dataset 7	0.604253665	0.604253665	0.60408096	0.60373555	0.604253665
Dataset 8	0.742411686	0.746978069	0.740033006	0.716711696	0.747395989
Dataset 9	0.446349147	0.472854102	0.441481554	0.463936732	0.572291667
Dataset 10	0.561393018	0.568147989	0.593233353	0.556736378	0.558860184

Datasets	XGboost Regressor AOA	XGboost Regressor GSA	XGboost Regressor PSO	XGboost Regressor COA	XGboost Regressor Hybrid AOA-COA
Dataset 1	0.87177735	0.863665125	0.871006208	0.859068149	0.859731056
Dataset 2	0.248624847	0.248624847	0.228905549	0.092954066	0.250282476
Dataset 3	0.923507569	0.427007748	0.923507569	0.923507569	0.935082113
Dataset 4	0.821497358	0.74176045	0.784193581	0.428571429	0.864230167
Dataset 5	0.779222143	0.765240814	0.75787426	0.755159019	0.827848165
Dataset 6	0.764021323	0.750283496	0.745331411	0.730223342	0.773925669
Dataset 7	0.666227876	0.649843148	0.655304724	0.635525866	0.66643998
Dataset 8	0.821332248	0.79783179	0.8053256	0.782188283	0.844905793
Dataset 9	0.242856669	0.189470716	0.202933173	0.236156565	0.496937046
Dataset 10	0.639336123	0.592270245	0.615564738	0.606116216	0.558860184

Datasets	Support Vector Regressor AOA	Support Vector Regressor GSA	Support Vector Regressor PSO	Support Vector Regressor COA	Support Vector Regressor Hybrid AOA-COA
Dataset 1	0.543635085	0.496076009	0.542841345	0.464563268	0.840080767
Dataset 2	0.521847697	0.583099642	0.563060622	0.568597423	0.614938389
Dataset 3	0.794984957	0.692848988	0.900081793	0.821019389	0.945390192
Dataset 4	0.859440249	0.894080332	0.882369665	0.428571429	0.883960475
Dataset 5	0.745101043	0.844826858	0.844760484	0.828599173	0.832965357
Dataset 6	0.741316399	0.75778568	0.758116415	0.746857921	0.782497818
Dataset 7	0.541691138	0.560400549	0.558386892	0.541832585	0.81816073
Dataset 8	0.534413864	0.559606431	0.556574343	0.432346784	0.695116845
Dataset 9	0.542619337	0.534375299	0.554996547	0.559871544	0.704887109
Dataset 10	0.661969335	0.656161903	0.64490698	0.649495874	0.666633162

Figure 5.2: Average R-squared results over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

Median of R-squared over 10 runs

Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.71088899	0.689649865	0.709896834	0.703823391	0.71088899
Dataset 2	0.641114674	0.641114674	0.641114674	0.633779875	0.641114674
Dataset 3	0.973687238	0.759613432	0.973687238	0.967262444	0.973687238
Dataset 4	0.961643273	0.916113436	0.933010705	0.917562301	0.955927269
Dataset 5	0.906302292	0.889952529	0.896321512	0.886741515	0.906302292
Dataset 6	0.843700625	0.824523132	0.827035961	0.817794158	0.839212938
Dataset 7	0.567920354	0.557425948	0.567920354	0.559690652	0.567920354
Dataset 8	0.446860296	0.437765838	0.443840518	0.433604976	0.446860296
Dataset 9	0.535509611	0.579531741	0.649882034	0.678553567	0.807697797
Dataset 10	0.661324103	0.688126152	0.728940045	0.735160269	0.66405052
RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.848070673	0.850044072	0.857208837	0.843151973	0.859798636
Dataset 2	0.497180947	0.499112706	0.499112706	0.468028002	0.499112706
Dataset 3	0.945395625	0.718467858	0.942515725	0.94321725	0.942646454
Dataset 4	0.898082869	0.915138172	0.924955182	0.89281965	0.930607318
Dataset 5	0.804389268	0.812421633	0.817731009	0.797578822	0.818196925
Dataset 6	0.751390368	0.737819309	0.750407008	0.738885798	0.750407008
Dataset 7	0.586613139	0.586030471	0.58224622	0.592167006	0.586030471
Dataset 8	0.792824537	0.793873067	0.791897478	0.773215567	0.795602391
Dataset 9	0.388255785	0.466606963	0.400348413	0.378590449	0.527104792
Dataset 10	0.738661135	0.749535795	0.732559366	0.74342026	0.671570658
Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.827273366	0.815243871	0.815490136	0.807325719	0.827273366
Dataset 2	0.632364832	0.632364832	0.632364832	0.631660693	0.632364832
Dataset 3	0.927844972	0.748931423	0.927844972	0.7969977	0.927844972
Dataset 4	0.882917466	0.882917466	0.882917466	0.808752399	0.883493282
Dataset 5	0.823610129	0.822322347	0.816928859	0.784401527	0.823610129
Dataset 6	0.672339753	0.672339753	0.672339753	0.672339753	0.672339753
Dataset 7	0.604253665	0.604253665	0.604253665	0.603677982	0.604253665
Dataset 8	0.742411686	0.744484211	0.734104515	0.720187528	0.750718857
Dataset 9	0.536401364	0.619624438	0.575351257	0.558714266	0.572291667
Dataset 10	0.626870084	0.592230527	0.630813241	0.602590297	0.558860184
XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.86605959	0.849388276	0.861495876	0.857821305	0.863422956
Dataset 2	0.248624847	0.248624847	0.248624847	0.051868671	0.250282476
Dataset 3	0.780231104	0.575687941	0.923507569	0.923507569	0.935082113
Dataset 4	0.821497358	0.744448889	0.780738621	0.428571429	0.865367945
Dataset 5	0.779222143	0.757448855	0.753432229	0.750270618	0.827848165
Dataset 6	0.764021323	0.764021323	0.745499268	0.726142796	0.773925669
Dataset 7	0.666227876	0.666227876	0.666227876	0.636015226	0.66643998
Dataset 8	0.821332248	0.798834855	0.802928318	0.780656534	0.844905793
Dataset 9	0.381413923	0.316924678	0.291132722	0.380603782	0.496937046
Dataset 10	0.717912445	0.681526974	0.646223788	0.64751032	0.558860184
Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.541441815	0.545097265	0.537082672	0.462150723	0.840744713
Dataset 2	0.521847697	0.583099642	0.555249192	0.569174417	0.614938389
Dataset 3	0.780231104	0.771499256	0.927998067	0.927872475	0.945390192
Dataset 4	0.860916322	0.899785975	0.87856672	0.428571429	0.882127645
Dataset 5	0.737161097	0.845866238	0.845210109	0.838281473	0.832965357
Dataset 6	0.737569985	0.757841409	0.760304131	0.747384257	0.782531031
Dataset 7	0.541691138	0.560400549	0.560400549	0.555366406	0.81816073
Dataset 8	0.534413864	0.559606431	0.558717485	0.437977873	0.698222383
Dataset 9	0.605760659	0.659129079	0.653863971	0.654698969	0.704887109
Dataset 10	0.706486992	0.714069971	0.726861873	0.708994114	0.666633162

Figure 5.3: Median of R-squared results over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

STD of R-squared over 10 runs

Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.000875293	0.004883684	0.000743437	0.012740428	0
Dataset 2	1.11022E-16	1.11022E-16	0.005096856	0.005696015	1.1102230E-16
Dataset 3	0	0.238726236	0.001975321	0.004775312	0
Dataset 4	0.005047962	0.022569677	0.011050213	0.031438433	0.018661548
Dataset 5	1.11022E-16	0.00966341	0.009125084	0.008865762	1.1102230E-16
Dataset 6	0.001346306	0.005426154	0.00817277	0.006418959	0.005804547
Dataset 7	1.11022E-16	0.005141188	0.007576815	0.009350761	1.1102230E-16
Dataset 8	5.55112E-17	0.005202677	0.002493823	0.004971969	5.5511151E-17
Dataset 9	0.354295067	0.232739521	0.2263074	0.22190944	1.11E-16
Dataset 10	0.156750116	0.204435662	0.193044451	0.222198065	0.220178075
RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.001974422	0.003948228	0.002779721	0.006559052	0.001377423
Dataset 2	0.018361295	0.002114784	0.002114784	0.0235322	0
Dataset 3	0.004164515	0.231942691	0.000685066	0.003972361	0
Dataset 4	0.017462033	0.004947665	0.002069719	0.012336692	0.003843588
Dataset 5	0.010456046	0.004535955	0.00471662	0.007926726	1.11E-16
Dataset 6	0.012258557	0.01057418	0.012082046	0.013002585	1.1102230E-16
Dataset 7	0.019808635	0.005991067	0.003104737	0.021243298	0
Dataset 8	0.006243616	0.008309782	0.005132063	0.011026294	1.11E-16
Dataset 9	0.479921692	0.49119625	0.461833142	0.444455081	0.004795642
Dataset 10	0.24462208	0.192081789	0.122123859	0.186995434	0.153201131
Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.000408602	0.007336454	0.0066079	0.008402922	1.11022E-16
Dataset 2	0	0	0	0.039689932	0
Dataset 3	2.22045E-16	0.32414598	0.039088845	0.058017562	2.22045E-16
Dataset 4	0.002775175	0.00028215	0.022269376	0.029713342	0.002815904
Dataset 5	1.11022E-16	0.002711133	0.015947756	0.002523941	1.1102230E-16
Dataset 6	1.11022E-16	1.11022E-16	1.11022E-16	1.11022E-16	1.1102230E-16
Dataset 7	0	0	0.000263811	0.000172705	0
Dataset 8	0.008307171	0.00305434	0.007650207	0.012752108	0.006645737
Dataset 9	0.30927484	0.334284727	0.223906308	0.281906866	1.11E-16
Dataset 10	0.206367415	0.198462814	0.188381282	0.192639211	0.195966579
XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.00102318	0.006869581	0.001280729	0.004648537	0.004521636
Dataset 2	2.77556E-17	2.77556E-17	0.059157895	0.065086055	0
Dataset 3	1.11022E-16	0.420228662	1.11022E-16	1.11022E-16	1.11022E-16
Dataset 4	0	0.03624337	0.027992537	0.428571429	0.003413333
Dataset 5	0	0.009868308	0.010871541	0.012610812	0
Dataset 6	0	0.021159408	0.018124372	0.018361763	1.11E-16
Dataset 7	1.11022E-16	0.025028085	0.021846303	0.026117109	0
Dataset 8	1.11022E-16	0.015699028	0.012135984	0.01511183	0.011954246
Dataset 9	0.531587134	0.558139222	0.564418497	0.60589524	5.55E-17
Dataset 10	0.208076799	0.24013293	0.207428763	0.199811022	0.195966579
Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.001790798	0.031609293	0.004564467	0.031155147	0.00199184
Dataset 2	1.11022E-16	0	0.013214839	0.014527578	0
Dataset 3	0.044261558	0.208034514	0.056101241	0.154167446	1.1102230E-16
Dataset 4	0.004464564	0.011150667	0.013918642	0.428571429	0.004351359
Dataset 5	0.023819841	0.003152047	0.00281832	0.014836982	0
Dataset 6	0.00482361	0.004496182	0.007580935	0.013530135	1.2643645E-04
Dataset 7	0	0	0.002466216	0.01975701	0
Dataset 8	0	1.11022E-16	0.005228632	0.049469447	0.009316615
Dataset 9	0.251044324	0.258460323	0.239218164	0.226599596	1.11E-16
Dataset 10	0.134428699	0.14118056	0.184070603	0.176760168	0.240683686

Figure 5.4: STD of R-squared results over 10 runs for the hybrid AOA-COA, AOA, PSO, GSA with different regressors.

Best R-squared over 10 runs

Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.71088899	0.708378607	0.71088899	0.710785988	0.71088899
Dataset 2	0.641114674	0.641114674	0.641114674	0.641114674	0.641114674
Dataset 3	0.973687238	0.945053136	0.973687238	0.972904469	0.973687238
Dataset 4	0.970326552	0.950445738	0.953242583	0.958122345	0.964908181
Dataset 5	0.906302292	0.90043538	0.906302292	0.899507458	0.906302292
Dataset 6	0.843700625	0.832273797	0.843700625	0.834505922	0.843700625
Dataset 7	0.567920354	0.567920354	0.567920354	0.567920354	0.567920354
Dataset 8	0.446860296	0.444500712	0.443911866	0.441040405	0.446860296
Dataset 9	0.835113122	0.759828826	0.829734781	0.864165092	0.807697797
Dataset 10	0.864189068	0.864189068	0.854053118	0.863962493	0.863962493
RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.859900085	0.855178856	0.859798636	0.85321965	0.859798636
Dataset 2	0.526509716	0.499112706	0.499112706	0.519697995	0.499112706
Dataset 3	0.952538293	0.884654041	0.942646454	0.948494374	0.942646454
Dataset 4	0.928145873	0.92635595	0.929221017	0.912114683	0.939166795
Dataset 5	0.81538003	0.818196925	0.818196925	0.813517323	0.818196925
Dataset 6	0.769086045	0.750407008	0.750407008	0.757729374	0.750407008
Dataset 7	0.621962398	0.586030471	0.586030471	0.627872974	0.586030471
Dataset 8	0.798466175	0.793873067	0.795602391	0.790193836	0.795602391
Dataset 9	0.820706407	0.829333866	0.846386375	0.841935475	0.536893854
Dataset 10	0.854699791	0.831771466	0.85855074	0.836339812	0.886516823
Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.827273366	0.827273366	0.827273366	0.827273366	0.827273366
Dataset 2	0.632364832	0.632364832	0.632364832	0.632364832	0.632364832
Dataset 3	0.927844972	0.927551064	0.927844972	0.927844972	0.927844972
Dataset 4	0.888963532	0.882917466	0.883493282	0.882917466	0.888963532
Dataset 5	0.823610129	0.823610129	0.823610129	0.790865756	0.823610129
Dataset 6	0.672339753	0.672339753	0.672339753	0.672339753	0.672339753
Dataset 7	0.604253665	0.604253665	0.604253665	0.604253665	0.604253665
Dataset 8	0.750718857	0.750718857	0.750718857	0.734104515	0.750718857
Dataset 9	0.744625759	0.744625759	0.696561025	0.777873718	0.572291667
Dataset 10	0.819993346	0.849351225	0.808	0.819993346	0.842989908
XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.86605959	0.863545684	0.866059592	0.869311252	0.863422956
Dataset 2	0.248624847	0.248624847	0.248624847	0.248624847	0.250282476
Dataset 3	0.923507569	0.734766427	0.923507569	0.923507569	0.935082113
Dataset 4	0.821497358	0.771994576	0.821497358	0.428571429	0.865367945
Dataset 5	0.779222143	0.779222143	0.779222143	0.779222143	0.827848165
Dataset 6	0.764021323	0.764021323	0.764021323	0.764021323	0.773925669
Dataset 7	0.666227876	0.666227876	0.666227876	0.666227876	0.66643998
Dataset 8	0.821332248	0.816549094	0.821332248	0.821332248	0.85686004
Dataset 9	0.85238224	0.842652568	0.858568147	0.842652568	0.496937046
Dataset 10	0.828849562	0.863077559	0.869818279	0.828835109	0.842989908
Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.541441815	0.545097265	0.545097265	0.545097265	0.840744713
Dataset 2	0.521847697	0.583099642	0.583099642	0.583099642	0.614938389
Dataset 3	0.927769631	0.878473632	0.928329236	0.928329236	0.945390192
Dataset 4	0.861532335	0.902287745	0.899785975	0.428571429	0.890133595
Dataset 5	0.816560566	0.848511074	0.848333366	0.840739072	0.832965357
Dataset 6	0.75101074	0.766751929	0.766751929	0.764559479	0.782654481
Dataset 7	0.541691138	0.560400549	0.560400549	0.560400549	0.81816073
Dataset 8	0.534413864	0.559606431	0.559606431	0.542440646	0.698222383
Dataset 9	0.866154812	0.828067143	0.824325471	0.824325471	0.704887109
Dataset 10	0.836219441	0.840802274	0.787184	0.840802274	0.879377754

Figure 5.5: Best R-squared results over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

Worst R-squared over 10 runs					
Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.71088899	0.634351212	0.703805332	0.667924917	0.71088899
Dataset 2	0.641114674	0.641114674	0.626593183	0.626226271	0.641114674
Dataset 3	0.973687238	0.350755073	0.96867913	0.955375339	0.973687238
Dataset 4	0.95293857	0.860965604	0.911614697	0.852226676	0.901852919
Dataset 5	0.906302292	0.872451289	0.876329452	0.868644223	0.906302292
Dataset 6	0.839212938	0.814936737	0.820950057	0.814596753	0.828381442
Dataset 7	0.567920354	0.557425948	0.543871109	0.543473909	0.567920354
Dataset 8	0.446860296	0.425431318	0.436954445	0.424951053	0.446860296
Dataset 9	-0.369339829	0.050409346	0.216303305	0.22348092	0.807697797
Dataset 10	0.261416957	0.213656548	0.155901552	0.080967462	0.080967462
RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.842876894	0.832663174	0.841543881	0.830331468	0.855178856
Dataset 2	0.477832998	0.493825745	0.493825745	0.455742327	0.499112706
Dataset 3	0.93905068	0.142991338	0.940241276	0.935481047	0.942646454
Dataset 4	0.87320096	0.907742418	0.922535797	0.872707917	0.926509885
Dataset 5	0.779887637	0.805848965	0.805848965	0.787026756	0.818196925
Dataset 6	0.724258824	0.723472789	0.723472789	0.715914265	0.750407008
Dataset 7	0.555134442	0.568355293	0.57855958	0.572464712	0.586030471
Dataset 8	0.775046337	0.771040919	0.781409114	0.74617457	0.795602391
Dataset 9	-0.841949429	-0.920370629	-0.770933943	-0.6933224	0.527104792
Dataset 10	0.102727227	0.17231894	0.416445716	0.1580823	0.300445152
Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.827273366	0.802423588	0.805955878	0.798519124	0.827273366
Dataset 2	0.632364832	0.632364832	0.632364832	0.531690959	0.632364832
Dataset 3	0.927844972	-0.227063625	0.7969977	0.7969977	0.927844972
Dataset 4	0.882917466	0.881976967	0.808752399	0.807831094	0.882917466
Dataset 5	0.823610129	0.813980707	0.785483119	0.78294373	0.823610129
Dataset 6	0.672339753	0.672339753	0.672339753	0.672339753	0.672339753
Dataset 7	0.604253665	0.604253665	0.603677982	0.603677982	0.604253665
Dataset 8	0.734104515	0.744484211	0.7331667	0.691547762	0.734104515
Dataset 9	-0.38672	-0.387596814	0.057348571	-0.137668571	0.572291667
Dataset 10	0.152288613	0.152288613	0.152288613	0.152288613	0.152288613
XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.86605959	0.839266066	0.85067505	0.852711576	0.854193206
Dataset 2	0.248624847	0.248624847	0.051431864	0.04939722	0.250282476
Dataset 3	0.923507569	-0.761515644	0.923507569	0.923507569	0.935082113
Dataset 4	0.821497358	0.64220711	0.718871933	0.428571429	0.853990169
Dataset 5	0.779222143	0.757448855	0.749535465	0.743470965	0.827848165
Dataset 6	0.764021323	0.711797689	0.726431794	0.710407077	0.773925669
Dataset 7	0.666227876	0.611612117	0.611612117	0.608936333	0.66643998
Dataset 8	0.821332248	0.774318628	0.783964311	0.759136663	0.832951547
Dataset 9	-0.648826005	-0.867471506	-0.869245775	-1.31536758	0.496937046
Dataset 10	0.085723834	0.085995859	0.085723834	0.085723834	0.152288613
Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.541441815	0.541441815	0.469364425	0.414438729	0.834105247
Dataset 2	0.521847697	0.583099642	0.549812146	0.551929975	0.614938389
Dataset 3	0.780231104	0.163833963	0.787879949	0.507861063	0.945390192
Dataset 4	0.846103305	0.870996713	0.864099409	0.428571429	0.876592122
Dataset 5	0.737161097	0.838925634	0.840142809	0.802913848	0.832965357
Dataset 6	0.737569985	0.750996407	0.741782798	0.72501751	0.78223007
Dataset 7	0.541691138	0.560400549	0.555366406	0.516928713	0.81816073
Dataset 8	0.534413864	0.559606431	0.542440646	0.377259345	0.667166999
Dataset 9	0.041873225	0.083608752	0.083608752	0.137919322	0.704887109
Dataset 10	0.392717079	0.401594596	0.202862958	0.202862958	0.013306311

Figure 5.6: Worst R-squared results over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

Average Feature Ratio over 10 runs					
Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.8	0.67	0.74	0.74	0.8
Dataset 2	0.33	0.33	0.33	0.35	0.33
Dataset 3	0.5	0.13	0.5	0.53	0.5
Dataset 4	0.53	0.48	0.43	0.57	0.49
Dataset 5	0.73	0.59	0.66	0.65	0.73
Dataset 6	0.55	0.26	0.42	0.44	0.48
Dataset 7	0.50	0.40	0.45	0.57	0.5
Dataset 8	0.60	0.55	0.54	0.54	0.6
Dataset 9	0.33	0.27	0.21	0.41	0.8
Dataset 10	0.47	0.43	0.41	0.40	0.41
RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor	RandomForest Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.53	0.49	0.5	0.54	0.49
Dataset 2	0.417	0.33	0.33	0.48	0.33
Dataset 3	0.46	0.13	0.38	0.48	0.38
Dataset 4	0.47	0.4	0.33	0.54	0.27
Dataset 5	0.38	0.34	0.31	0.55	0.27
Dataset 6	0.20	0.19	0.16	0.45	0.18
Dataset 7	0.63	0.52	0.60	0.38	0.5
Dataset 8	0.26	0.19	0.37	0.44	0.2
Dataset 9	0.54	0.43	0.54	0.49	0.28
Dataset 10	0.45	0.43	0.33	0.23	0.27
Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor	Kneighbors Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.48	0.37	0.4	0.49	0.47
Dataset 2	0.17	0.17	0.17	0.53	0.17
Dataset 3	0.2	0.13	0.25	0.36	0.175
Dataset 4	0.3	0.15	0.33	0.43	0.18
Dataset 5	0.25	0.14	0.26	0.35	0.19
Dataset 6	0.55	0.48	0.50	0.55	0.53
Dataset 7	0.17	0.17	0.20	0.28	0.17
Dataset 8	0.36	0.14	0.41	0.54	0.27
Dataset 9	0.60	0.48	0.46	0.13	0.17
Dataset 10	0.46	0.38	0.24	0.22	0.26
XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor	XGboost Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.8	0.55	0.63	0.6	0.38
Dataset 2	0.17	0.17	0.2	0.35	0.17
Dataset 3	0.66	0.13	0.64	0.73	0.54
Dataset 4	0.31	0.38	0.33	0.43	0.18
Dataset 5	0.42	0.23	0.32	0.40	0.35
Dataset 6	0.28	0.27	0.36	0.49	0.28
Dataset 7	0.50	0.45	0.47	0.68	0.50
Dataset 8	0.30	0.30	0.33	0.47	0.40
Dataset 9	0.48	0.40	0.39	0.30	0.44
Dataset 10	0.50	0.34	0.47	0.45	0.26
Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor	Support Vector Regressor
	AOA	GSA	PSO	COA	Hybrid AOA-COA
Dataset 1	0.1	0.1	0.12	0.37	0.4
Dataset 2	0.33	0.17	0.3	0.35	0.83
Dataset 3	0.25	0.13	0.14	0.25	0.5
Dataset 4	0.45	0.07	0.14	0.43	0.25
Dataset 5	0.31	0.29	0.35	0.37	0.73
Dataset 6	0.35	0.29	0.33	0.48	0.48
Dataset 7	0.17	0.17	0.17	0.35	0.33
Dataset 8	0.10	0.10	0.10	0.33	0.21
Dataset 9	0.50	0.43	0.39	0.40	0.20
Dataset 10	0.54	0.37	0.35	0.21	0.30

Figure 5.7: Average feature ratio over 10 runs for AOA, PSO, GSA, COA, and hybrid AOA-COA with different regressors.

The following tables show the ranking among AOA, PSO, GSA, COA, and hybrid AOA-COA in each regressor and through the 6 evaluation criteria. When considering having the lower feature ratio and higher R-squared together, the hybrid AOA-COA algorithm performed best when combined with XGBoost, Random Forest, and K-Neighbor regressors across all 10 datasets. Due to the high-dimensional nature of climate-related issues, which involve multiple contributing factors, the best-selected combined features can differ based on the specific characteristics of the dataset. This can lead to varying and potentially higher feature ratios. A higher feature ratio indicates that the model needs more features to achieve more accurate performance.

Linear Regression	**Ranking**
(AOA)	1
Hybrid AOA-COA	2
(PSO)	3
(GSA)	4
(COA)	5

Table 5.8: Hybrid AOA-COA Ranking Over All Evaluation Criteria with Linear Regression

RandomForest Regressor	**Ranking**
Hybrid AOA-COA	1
(PSO)	2
(AOA)	3
(GSA)	4
(COA)	5

Table 5.9: Hybrid AOA-COA Ranking Over All Evaluation Criteria with Random Forest Regressor

K-Neighbors Regressor	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(GSA)	3
(PSO)	4
(COA)	5

Table 5.10: Hybrid AOA-COA Ranking Over All Evaluation Criteria with K-Neighbors Regressor

XGBoost Regressor	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(PSO)	3
(GSA)	4
(COA)	5

Table 5.11: Hybrid AOA-COA Ranking Over All Evaluation Criteria with XGBoost Regressor

Support Vector Regressor	**Ranking**
Hybrid AOA-COA	1
(GSA)	2
(PSO)	3
(AOA)	4
(COA)	5

Table 5.12: Hybrid AOA-COA Ranking Over All Evaluation Criteria with Support Vector Regressor

In general, as shown in table 5.13 the hybrid AOA-COA algorithm is the best over all 6 evaluation metrics and across all regressors.

Over All Regressors	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(PSO)	3
(GSA)	4
(COA)	5

Table 5.13: Hybrid AOA-COA Ranking Over All Evaluation Metrics and All Regressors

5.5 MSE Results for Reconfirmation

In addition to evaluating the performance of the proposed hybrid AOA-COA model using the six mentioned evaluation criteria, We compared the hybrid model performance with the initial 11 algorithms in combination with the 5 regressors using average Mean Squared Error (MSE) results over 10 runs and across the 10 datasets (in the same situation) in this stage.

The results affirmed previous findings, indicating our hybrid AOA-COA model is the best option, ranking highest among all regressors across all datasets. The MSE results and the ranking are shown in the following tables.

Linear Regression	GA	AOA	GOA	GSA	HHO	GWO	WOA	PSO	COA	DBO	POA	Hybrid AOA-COA
Dataset 1	0.011700935	**0.009334321**	0.00939977	0.00971666	0.0250979	0.009682	0.019204649	0.009547618	0.009595617	0.010334865	0.010092823	**0.009334321**
Dataset 2	0.059346404	**0.051946537**	0.05258706	0.05194654	0.0742998	0.078656	0.063948264	0.052413399	0.053008775	0.052712046	0.052503211	**0.051946537**
Dataset 3	0.004376659	**0.003416395**	0.00370436	0.00378414	0.0366141	0.008328	0.021891114	0.003574242	0.004425247	0.004413671	0.004198912	**0.003416395**
Dataset 4	0.003406391	**0.001168982**	0.0018109	0.00265222	0.0192055	0.009555	0.006948804	0.00195003	0.002638345	0.002388106	0.003076603	0.001497252
Dataset 5	0.007287647	**0.005756047**	0.006087	0.00690069	0.0210499	0.009743	0.010180298	0.006524931	0.006943962	0.006677062	0.006653874	**0.005756047**
Dataset 6	0.034270256	0.025094036	0.02786919	0.02831124	0.0438465	**0.022936**	0.041587283	0.027249183	0.028676996	0.030176972	0.02962618	0.026117153
Dataset 7	0.013821703	0.012859651	0.01309986	0.01304705	0.0117331	**0.008472**	0.038545929	0.013011448	0.013213793	0.013158872	0.01321714	0.012859651
Dataset 8	0.019372798	0.018692685	0.01895944	0.01899462	0.018417	**0.016681**	0.04364384	0.018858581	0.019122518	0.019098669	0.019204397	0.018692685
Dataset 9	0.028737081	0.031494881	0.0376536	0.02979467	0.0296984	0.029448	0.028423642	0.026887455	0.022089861	0.022103718	0.021982545	**0.009968946**
Dataset 10	0.022529443	0.02278958	0.02405659	0.0230925	0.0237736	0.022195	0.022875119	0.021031669	0.020323615	**0.019674745**	0.019857688	0.020518639
Sum of Ranks	90	35	61	67	95	66	103	43	67	65	64	24
Final Average Ranking	**10**	**2**	**4**	**8.5**	**11**	**7**	**12**	**3**	**8.5**	**6**	**5**	**1**

Random Forest Regressor	GA	AOA	GOA	GSA	HHO	GWO	WOA	PSO	COA	DBO	POA	Hybrid AOA-COA
Dataset 1	0.005024423	0.004679987	0.00471099	0.00498447	0.0218025	0.006684	0.00945157	0.004610503	0.005147843	0.004994044	0.005026347	**0.004543305**
Dataset 2	0.077174549	0.072430488	0.07417779	0.07265354	0.1338587	0.164369	0.085173481	0.072653537	0.075469516	0.074481706	**0.072425579**	0.072500485
Dataset 3	0.007860052	0.007183148	0.00765547	0.00752619	0.0320099	0.014105	0.011682066	0.007492721	0.007449982	**0.007136501**	0.007386262	0.007446666
Dataset 4	0.002943729	0.003007573	0.00229914	0.00251229	0.0440399	0.013973	0.00377996	0.002214216	0.003142939	0.003231731	0.003203805	**0.002036751**
Dataset 5	0.011944725	0.012094535	0.01159052	0.01152318	0.0297256	0.015998	0.013843778	0.01137045	0.012476759	0.012330855	0.012475507	**0.011168544**
Dataset 6	0.043735334	0.039796931	0.04391465	0.04215902	0.0547789	**0.036068**	0.048056857	0.041223054	0.042222453	0.042400625	0.04203646	0.039957701
Dataset 7	0.013009586	0.012329642	0.01254035	0.01244245	0.0157276	**0.007391**	0.013612164	0.012421147	0.599555495	0.011952947	0.012149547	0.012320654
Dataset 8	0.015507942	0.007035902	0.00733352	0.00714201	0.0253334	0.009527	0.012547916	0.007102418	0.007737623	0.007702115	0.007442114	**0.006907369**
Dataset 9	0.036326764	0.034816795	0.039141	0.03270909	0.0376213	0.036117	0.036450658	0.035417579	0.032842655	0.033374664	0.034530766	**0.024311902**
Dataset 10	0.018417855	0.021156039	0.02358927	0.02055214	0.0191063	0.020066	**0.018017885**	0.018265899	0.018939712	0.018782701	0.019783857	0.020188724
Sum of Ranks	76	45	72	52.5	111	82	93	38.5	73	55	54	28
Final Average Ranking	9	3	7	4	12	10	11	2	8	6	5	**1**

K-Neighbors Regressor	GA	AOA	GOA	GSA	HHO	GWO	WOA	PSO	COA	DBO	POA	Hybrid AOA-COA
Dataset 1	0.007402238	**0.005576702**	0.00594142	0.0060098	0.0258885	0.009748	0.010372527	0.006005592	0.006253316	0.006184511	0.006401947	**0.005576702**
Dataset 2	0.061596235	**0.053213025**	0.05715031	0.05321303	0.0728411	0.085274	0.067748851	0.053213025	0.056657944	0.058604899	0.054735872	**0.053213025**
Dataset 3	0.021739104	**0.00936846**	0.01876495	0.00936846	0.032654	0.01551	0.030096563	0.011149488	0.020245302	0.017011782	0.020525886	**0.00936846**
Dataset 4	0.005732229	0.003418343	0.00547291	0.00348851	0.0148931	0.011792	0.006081829	0.0037048	0.005254857	0.005696457	0.0052576	**0.003411829**
Dataset 5	0.023813156	**0.010836**	0.01246978	0.01098058	0.0245013	0.01295	0.013196	0.011834711	0.013205511	0.013007067	0.013246533	**0.010836**
Dataset 6	0.0524556	0.0524556	0.0524556	0.0524556	0.0532458	**0.032841**	0.05695794	0.0524556	0.0524556	0.0524556	0.0524556	0.0524556
Dataset 7	0.01312617	0.011778291	0.01197323	0.01177829	0.0128459	**0.008428**	0.015688152	0.011783431	0.011793711	0.011793711	0.011790285	0.011778291
Dataset 8	0.010229708	0.008704885	0.00907708	0.00855057	0.0237126	0.012302	0.01204359	0.008785269	0.009573385	0.009202785	0.009520331	**0.008536446**
Dataset 9	0.030195936	0.031420224	0.03345322	0.02922156	0.0310934	0.030107	0.029890512	0.033618888	0.030375432	0.030328344	0.030106944	**0.0221724**
Dataset 10	0.027759672	0.02681316	0.02739355	0.02653906	0.0269896	0.027996	0.026874936	**0.024957936**	0.027668952	0.027482328	0.027482328	0.027439344
Sum of Ranks	91	34.5	66	30.5	106	73.5	91	46.5	74.5	70	70	26.5
Final Average Ranking	10.5	3	5	2	12	8	10.5	4	9	6.5	6.5	**1**

XGBoost Regressor	GA	AOA	GOA	GSA	HHO	GWO	WOA	PSO	COA	DBO	POA	Hybrid AOA-COA
Dataset 1	0.005135408	**0.004324439**	0.00454313	0.00483466	0.0243783	0.008416	0.005105078	0.004510557	0.004840715	0.004931905	0.004804141	0.004528764
Dataset 2	0.137553762	0.108757127	0.12701482	0.10875713	0.1635787	0.159809	0.150195065	0.11161138	0.131289556	0.128214387	0.129832658	**0.108517195**
Dataset 3	0.017856835	0.00993162	0.00993162	0.00993162	0.0590152	0.017523	0.00993162	0.00993162	0.00993162	0.00993162	0.00993162	**0.008428805**
Dataset 4	0.007264117	0.005314279	0.00754869	0.00768816	0.0415097	0.018775	0.01173564	0.006424865	0.009084626	0.00945079	0.007898034	**0.004042062**
Dataset 5	0.016272628	0.013562847	0.01504687	0.01442175	0.021644	0.020219	0.017301614	0.014874292	0.015041095	0.014950359	0.015017537	**0.010575649**
Dataset 6	0.043226426	0.037778166	0.04241591	0.03997747	0.0658663	0.048072	0.046050882	0.04077026	0.043188933	0.044515611	0.043037293	**0.036192565**
Dataset 7	0.011257628	0.009933801	0.01047992	0.01042145	0.018713	0.016055	0.01301679	0.010258898	0.010847561	0.011558725	0.011643536	**0.009927488**
Dataset 8	0.007310026	0.00603786	0.00700297	0.00683203	0.0293468	0.009004	0.008300983	0.006578785	0.007360683	0.007343699	0.007143638	**0.005241221**
Dataset 9	0.044988226	0.039405465	0.04785596	0.04447854	0.0472946	0.041782	0.047782903	0.043794444	0.039333449	0.039913993	0.041831078	**0.026078784**
Dataset 10	0.023905139	**0.022372288**	0.02905503	0.02466388	0.028331	0.024769	0.027514078	0.023841628	0.024708632	0.023296856	0.024797181	0.024367293
Sum of Ranks	78	23	66.5	48	117	100	95.5	38.5	66.5	64.5	66.5	16
Final Average Ranking	9	2	7	4	12	11	10	3	7	5	7	**1**

Support Vector Regressor	GA	AOA	GOA	GSA	HHO	GWO	WOA	PSO	COA	DBO	POA	Hybrid AOA-COA
Dataset 1	0.019887784	0.014805142	0.01694681	0.01469892	0.0182365	0.011246	0.019245332	0.015042562	0.017706311	0.017086234	0.016138184	**0.005163199**
Dataset 2	0.080476157	0.069209729	0.06155856	0.06034387	0.1176973	0.091776	0.112110338	0.063244401	0.062442982	0.061959902	0.062844838	**0.0557354**
Dataset 3	0.048347931	0.02661873	0.02685218	0.00931151	0.0301893	0.0137	0.08560502	0.012973174	0.023238473	0.025134423	0.032308165	**0.007090425**
Dataset 4	0.00555584	0.004184665	0.00426904	**0.00315338**	0.013761	0.010475	0.01477983	0.003502023	0.005031822	0.004389405	0.004156638	0.003454662
Dataset 5	0.022289842	0.015658978	0.00966665	**0.00953261**	0.0114465	0.011816	0.052089857	0.009536689	0.010529513	0.011506959	0.010097585	0.010261289
Dataset 6	0.044684307	0.041413029	0.03860532	0.03877644	0.0507045	**0.03453**	0.104786585	0.038723491	0.040525879	0.040902014	0.042306318	0.034820237
Dataset 7	0.014690731	0.013640291	0.01402374	0.01308346	0.0124752	0.007153	0.015245156	0.013143388	0.013636081	0.013694209	0.013280273	**0.005411941**
Dataset 8	0.022982631	0.015733919	0.02010652	0.01488257	0.0213877	0.01275	0.023431552	0.014985032	0.019183152	0.019095942	0.019646697	**0.010303156**
Dataset 9	0.025726155	0.025477478	0.02792475	0.02617322	0.0275631	0.027561	0.025954507	0.025078041	0.025111209	0.02511129	0.025108824	**0.015298652**
Dataset 10	0.021489839	0.020861097	0.02241908	0.02114656	0.023404	0.020884	0.022137802	0.021495651	0.021569466	0.021569466	0.021545142	**0.019903782**
Sum of Ranks	96	63	72	34	96	53	112	41	65.5	68.5	63	16
Final Average Ranking	10.5	5.5	9	2	10.5	4	12	3	7	8	5.5	**1**

Figure 5.8: MSE results with Final Ranking for hybrid AOA-COA and all 11 algorithms in each regressor.

Across all regressors	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(PSO)	3
(GSA)	4
(POA)	5
(GOA)	6
(DBO)	7
(COA)	8
(GWO)	9
(GA)	10
(WOA)	11
(HHO)	12

Table 5.14: MSE Ranking for hybrid AOA-COA and all other algorithms over all regressors.

In the table 5.15, we have the average of MSE results over 10 runs across all datasets and Regressors.The results show that the proposed hybrid AOA-COA algorithm achieved a 13% improvement across all datasets and regressors, outperforming the top three algorithms in the experiments due to its exceptional feature selection optimization.

Algorithm	**Average Mean MSE**
AOA	0.022
GSA	0.022
PSO	0.022
COA	0.035
Hybrid AOA-COA	0.019

Table 5.15: Average of Mean MSE Over 10 Runs Across All Datasets and All Regressors

5.6 Dataset 1 as an Example

In this section, we have a table displaying the R-squared values of the regressor models without feature selection for Dataset 1; we gave "CO2" as the independent variable and "Temp" as the target variable to the machine. As indicated in the table, the R-squared for the Linear regression, Random Forest regressor, K-Neighbor regressor, XGBoost regressor, and Support Vector regressor are 0.5425, 0.4922, 0.555, 0.3452, 0.6084 respectively.

CO2 - Temp	**R-squared**
Linear Regression	0.5425
Random Forest Regressor	0.4922
K-Neighbors Regressor	0.555
XGBoost Regressor	0.3452
Support Vector Regressor	0.6084

Table 5.16: R-squared values for different regressors for the CO2 - Temp relationship.

The goal is to understand better how the combination of features can significantly impact the target variable, particularly in the context of the complex issue of climate. As mentioned earlier, the number of subsets in a feature selection problem with "n" features is 2 to the power of "n," making it a complex and challenging situation. However, as demonstrated, natural-inspired algorithms can be highly beneficial and practical in feature selection optimization.

Therefore, we also compiled tables displaying the results of each evaluation metric for AOA, PSO, GSA, COA, and the hybrid AOA-COA approach using Dataset 1 for comparison. The R-squared results for the hybrid AOA-COA, combined with the five mentioned regressors, in the same order: 0.7108, 0.8592, 0.8272, 0.8597, and 0.840 showing the effectiveness of feature selection. The results indicate that the proposed hybrid AOA-COA achieved the highest rank in 5 out of 7 evaluation metrics using Dataset 1.

Average R-squared over 10 runs - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.710563296	0.705707823	0.710517272	0.69873803	0.71088899
RandomForest Regressor	0.858748098	0.851067656	0.854230503	0.844597571	0.859280676
Kneighbors Regressor	0.827137165	0.812486138	0.827273366	0.809426421	0.827273366
XGboost Regressor	0.87177735	0.863665125	0.871006208	0.859068149	0.859731056
Support Vector Regressor	0.543635085	0.496076009	0.542841345	0.464563268	0.840080767
Sum of Ranks	10	19	12.5	25	8.5
Final Ranking	2	4	3	5	**1**

Median R-squared over 10 runs - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.71088899	0.689649865	0.709896834	0.703823391	0.71088899
RandomForest Regressor	0.848070673	0.850044072	0.857208837	0.843151973	0.859798636
Kneighbors Regressor	0.827273366	0.815243871	0.815490136	0.807325719	0.827273366
XGboost Regressor	0.86605959	0.849388276	0.861495876	0.857821305	0.863422956
Support Vector Regressor	0.541441815	0.545097265	0.537082672	0.462150723	0.840744713
Sum of Ranks	10.5	19	15	23	7.5
Final Ranking	2	4	3	5	**1**

STD of R-squared over 10 runs - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.000875293	0.004883684	0.000743437	0.012740428	0
RandomForest Regressor	0.001974422	0.003948228	0.002779721	0.006559052	0.001377423
Kneighbors Regressor	0.000408602	0.007336454	0.0066079	0.008402922	1.11022E-16
XGboost Regressor	0.00102318	0.006869581	0.001280729	0.004648537	0.004521636
Support Vector Regressor	0.001790798	0.031609293	0.004564467	0.031155147	0.00199184
Sum of Ranks	9	22	13	23	8
Final Ranking	2	4	3	5	**1**

Best R-squared over 10 runs - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.71088899	0.708378607	0.71088899	0.710785988	0.71088899
RandomForest Regressor	0.859900085	0.855178856	0.859798636	0.85321965	0.859798636
Kneighbors Regressor	0.827273366	0.827273366	0.827273366	0.827273366	0.827273366
XGboost Regressor	0.86605959	0.863545684	0.866059592	0.869311252	0.863422956
Support Vector Regressor	0.541441815	0.545097265	0.545097265	0.545097265	0.840744713
Sum of Ranks	13.5	19	12	16	14.5
Final Ranking	2	5	**1**	4	3

Figure 5.9: The results of each evaluation metric for AOA, PSO, GSA, COA, and hybrid AOA-COA in each regressor using Dataset 1

Worst R-squared over 10 runs - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.71088899	0.634351212	0.703805332	0.667924917	0.71088899
RandomForest Regressor	0.842876894	0.832663174	0.841543881	0.830331468	0.855178856
Kneighbors Regressor	0.827273366	0.802423588	0.805955878	0.798519124	0.827273366
XGboost Regressor	0.86605959	0.839266066	0.85067505	0.852711576	0.854193206
Support Vector Regressor	0.541441815	0.541441815	0.469364425	0.414438729	0.834105247
Sum of Ranks	8	20.5	17	22	7.5
Final Ranking	2	4	3	5	**1**

Average Feature Ratio over 10 runs- Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.8	0.67	0.74	0.74	0.8
RandomForest Regressor	0.53	0.49	0.5	0.54	0.49
Kneighbors Regressor	0.48	0.37	0.4	0.49	0.47
XGboost Regressor	0.8	0.55	0.63	0.6	0.38
Support Vector Regressor	0.1	0.1	0.12	0.37	0.4
Sum of Ranks	19	6.5	14.5	19.5	15.5
Final Ranking	4	**1**	2	5	3

MSE - Dataset 1					
Regressors \ Algorithms	**AOA**	**GSA**	**PSO**	**COA**	**Hybrid AOA-COA**
Linear Regression	0.009334321	0.009716664	0.009547618	0.009595617	0.009334321
RandomForest Regressor	0.004679987	0.004984471	0.004610503	0.005147843	0.004543305
Kneighbors Regressor	0.005576702	0.006009796	0.006005592	0.006253316	0.005576702
XGboost Regressor	0.004324439	0.004834657	0.004510557	0.004840715	0.004528764
Support Vector Regressor	0.014805142	0.014698923	0.015042562	0.017706311	0.005163199
Sum of Ranks	9.5	19	14	24	8.5
Final Ranking	2	4	3	5	**1**

Figure 5.10: The results of each evaluation metric for AOA, PSO, GSA, COA, and hybrid AOA-COA in each regressor using Dataset 1

5.6.1 Visualization

To illustrate how well the hybrid AOA-COA model performs, the scatter plots below compare the actual target values (y-test) with the predicted values (y-pred) using Dataset 1. Each point on the scatter plot corresponds to a single observation from the test set. The X-coordinate of the point represents the actual value, and the Y-coordinate represents the predicted value for that same observation.

The red diagonal line is a reference line where the predicted values equal the actual values (y-pred = y-test). If the model were perfect, all the points would lie exactly on this line. If the points cluster closely around the red diagonal line, the model's predictions are close to the actual values, indicating a good fit. If the points are widely scattered away from the diagonal line, the model's predictions deviate significantly from the actual values, indicating a poor fit. The closer the points are to the diagonal line, the better the model's performance.

The hybrid AOA-COA method, combined with the XGBoost regressor, identified CO2, N2O, CFC-11, and aerosols as the features with the most significant combined impact on the target variable "Temp." This model achieved higher accuracy than other regressors in the Dataset 1, as demonstrated by the scatter plot.

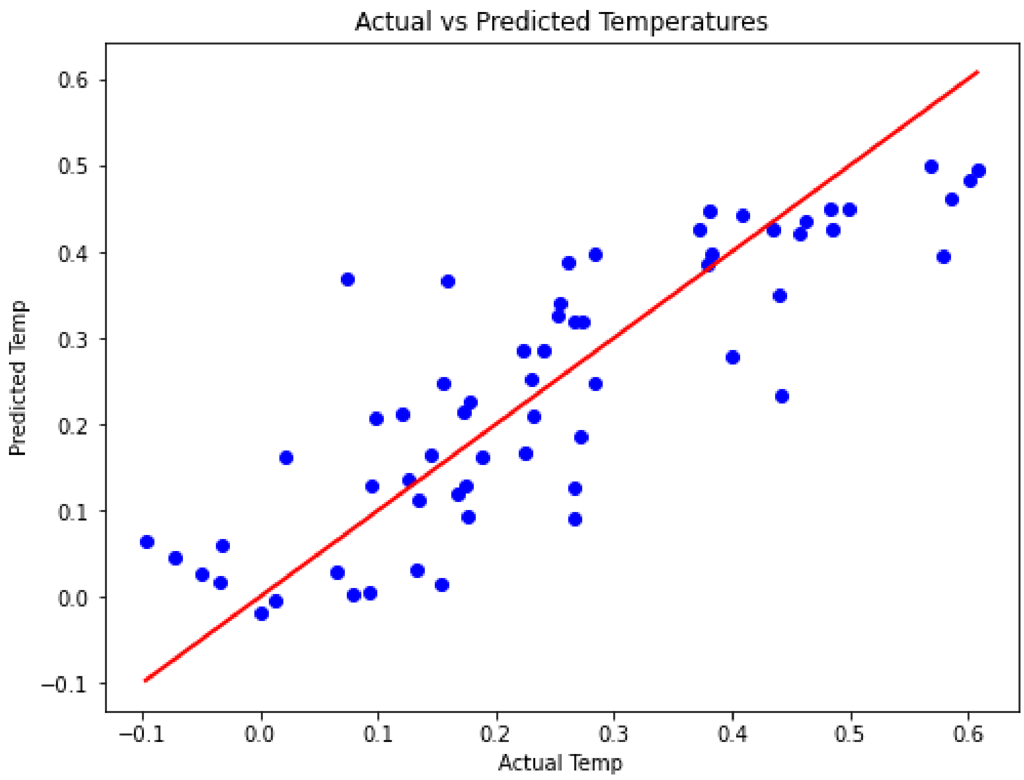

Figure 5.11: Scatter plot of actual vs predicted values-Linear regression-Dataset 1- for hybrid AOA-COA.

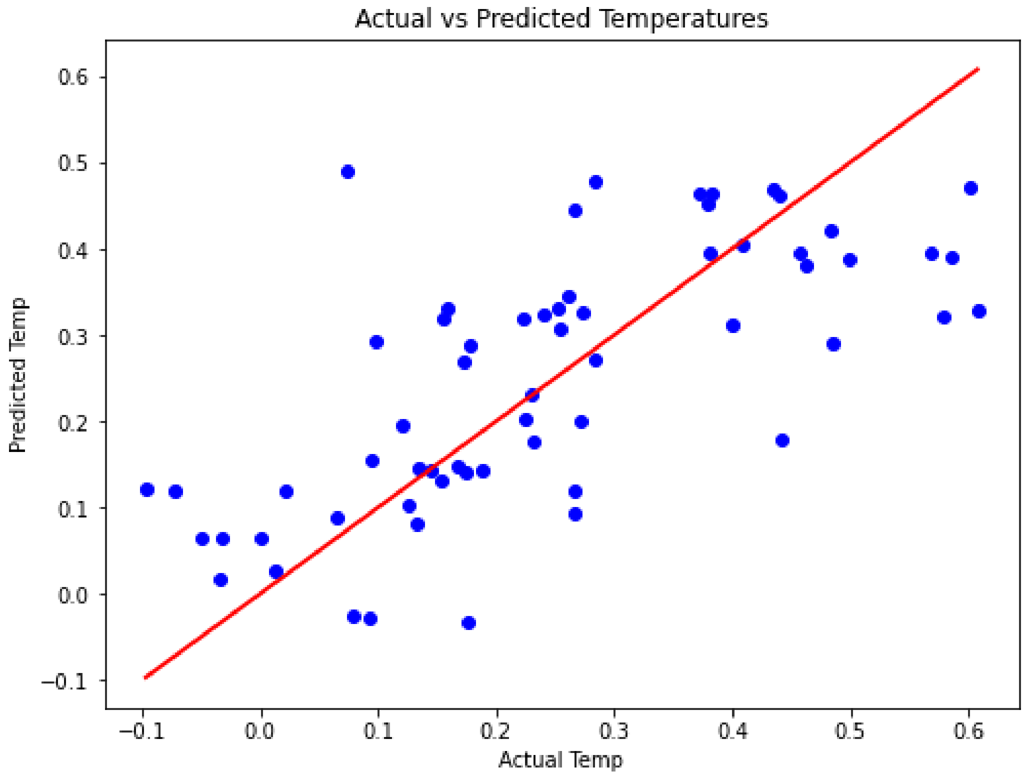

Figure 5.12: Scatter plot of actual vs predicted values-Random Forest Regressor- Dataset 1- for hybrid AOA-COA.

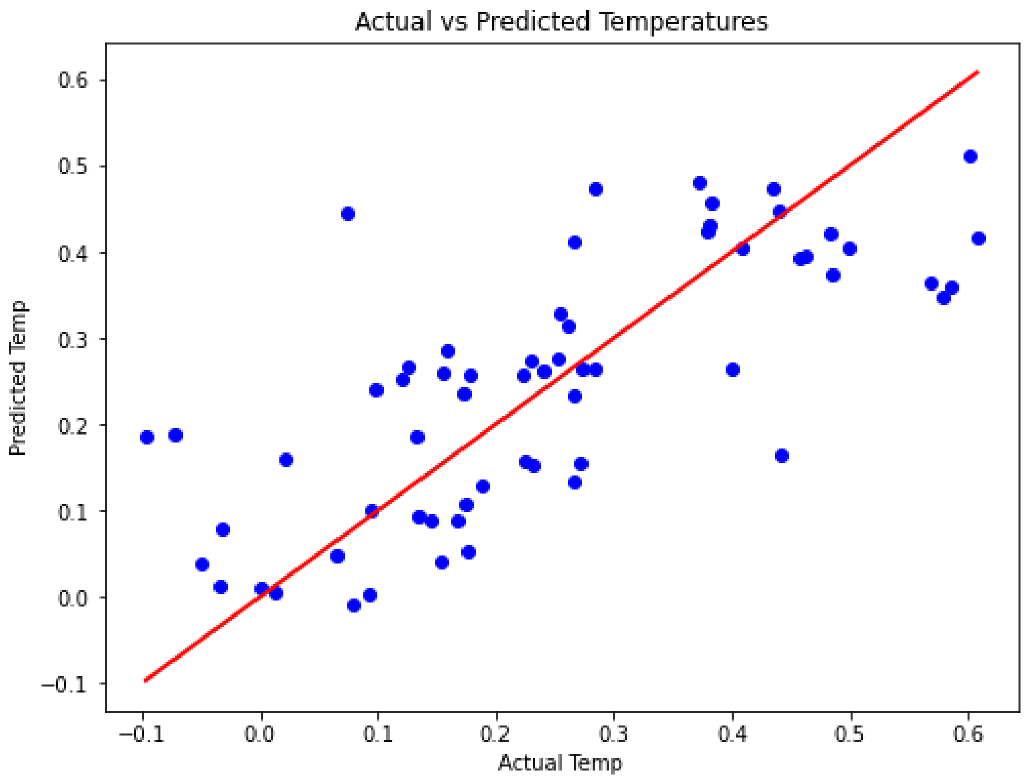

Figure 5.13: Scatter plot of actual vs predicted values-K-Neighbor Regressor-Dataset 1- for hybrid AOA-COA.

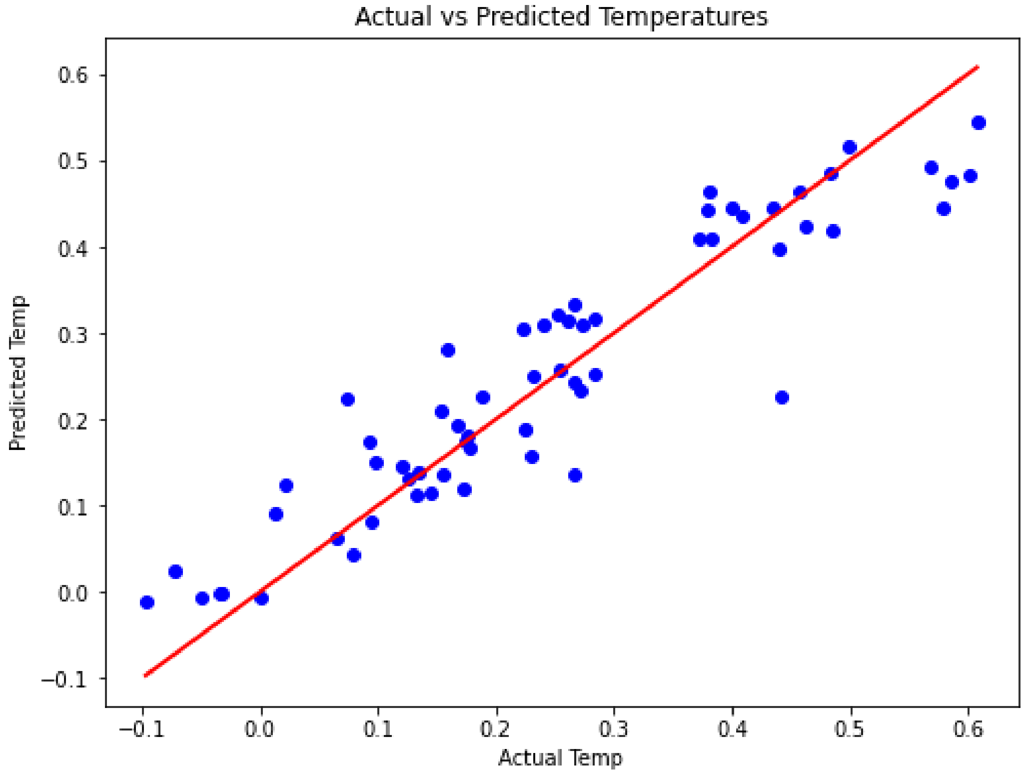

Figure 5.14: Scatter plot of actual vs predicted values-XGBoost Regressor-Dataset 1- for hybrid AOA-COA.

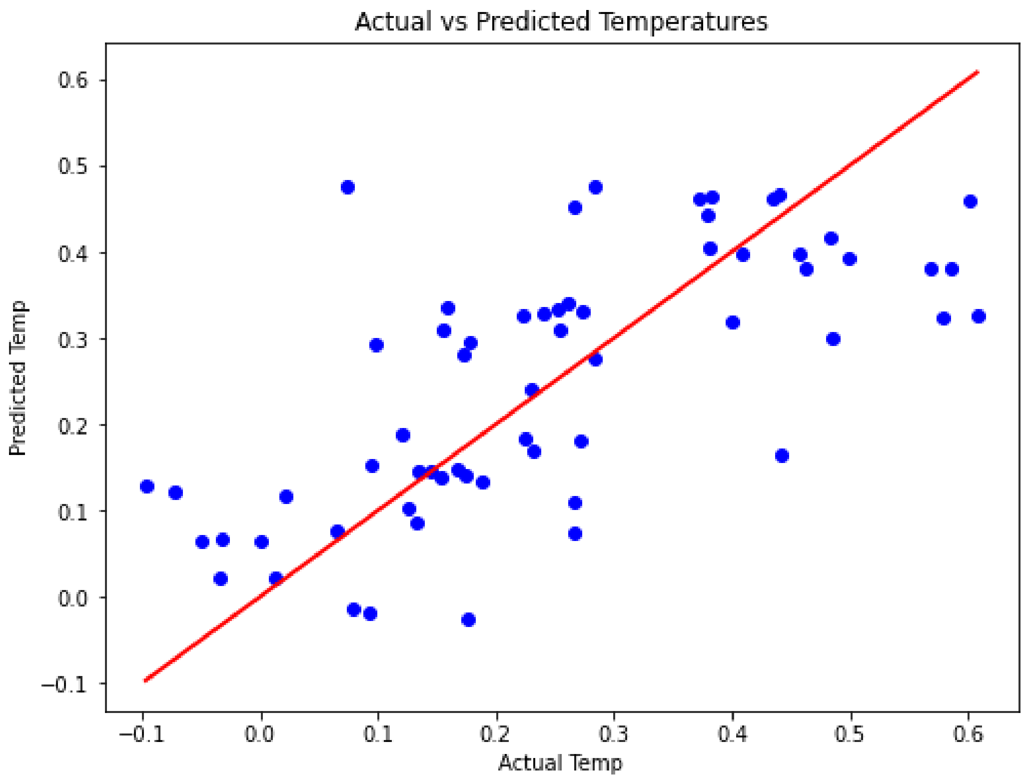

Figure 5.15: Scatter plot of actual vs predicted values-Support Vector Regressor- Dataset 1- for hybrid AOA-COA.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this research, we proposed and evaluated a hybrid swarm intelligence algorithm, AOA-COA, for feature selection in predicting global temperature changes. The hybrid algorithm demonstrated good improvements over individual nature-inspired algorithms, such as the Arithmetic Optimization Algorithm (AOA), Particle Swarm Optimization (PSO), and Gravitational Search Algorithm (GSA), which were identified as the top-performing individual algorithms in earlier evaluations within this study (Table 6.1, Table 6.2).

The proposed hybrid AOA-COA consistently ranked highest in terms of various performance metrics across multiple machine learning regressors and datasets. The hybrid model was tested across different regressors, including Linear regression, Random Forest regressor, K-Neighbors regressor, XGBoost regressor, and Support Vector regressor. The proposed hybrid algorithm outperformed the individual methods, validating its potential for practical applications in climate science models.

Across all Regressors	**Ranking**
(AOA)	1
(PSO)	2
(GSA)	3
(GOA)	4
(POA)	5
(DBO)	6
(COA)	7
(GA)	8
(GWO)	9
(WOA)	10
(HHO)	11

Table 6.1: Ranking of 11 algorithms across all regressors and 6 evaluation criteria (Final Ranking)

Over All Regressors	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(PSO)	3
(GSA)	4
(COA)	5

Table 6.2: Hybrid AOA-COA ranking across 6 evaluation criteria and over all regressors.

The final results revealed that the hybrid AOA-COA approach outperformed other models, achieving the best MSE and R-squared values among all regressors and across all 10 datasets. Also, in terms of feature ratio, it ranked first in performance with the Random Forest regressor and second with both the K-Neighbors and XGBoost regressors, confirming its robustness in handling high-dimensional climate datasets.

Moreover, the findings indicate that the hybrid AOA-COA, in combination with the Support Vector regressor, outperformed surprisingly other algorithm combinations with the Support Vector regressor.

Table 6.3, which ranks the algorithms based on MSE, demonstrated that the hybrid AOA-COA algorithm consistently achieved higher rankings compared to all other natural-inspired algorithms examined in this study. These results, combined with reconfirmation across 10 datasets and 10 independent runs, further solidified the conclusion that the hybrid AOA-COA model optimizes feature selection and enhances predictive accuracy when dealing with complex environmental data.

MSE employed as an additional evaluation metric to reconfirm the findings across six other performance criteria including Average R-squared, Median, Standard Deviation (STD), Best, Worst of R-squared, and Feature Ratio. The findings indicate that the suggested AOA-COA hybrid algorithm attained a 13% enhancement of MSE across all datasets and regressors, surpassing the top three algorithms in the experiments due to its outstanding optimization of feature selection. The consistent performance of the hybrid AOA-COA algorithm across these diverse metrics underscores its reliability and robustness for feature selection in global temperature prediction models.

Across all regressors	**Ranking**
Hybrid AOA-COA	1
(AOA)	2
(PSO)	3
(GSA)	4
(POA)	5
(GOA)	6
(DBO)	7
(COA)	8
(GWO)	9
(GA)	10
(WOA)	11
(HHO)	12

Table 6.3: MSE ranking for hybrid AOA-COA and all other algorithms over all regressors.

The findings from this study highlight the importance of hybrid approaches in feature selection, particularly when dealing with high-dimensional datasets. The combination of AOA as global search method with COA as local search refinements in the feature selection process enabled the model to reduce dimensionality effectively while preserving high predictive power. When examining the combination of a lower feature ratio and a higher R-squared, the hybrid AOA-COA algorithm showed the best performance when used with XGBoost, Random Forest, and K-Neighbor regressors across all ten datasets.

This research demonstrates the potential of hybrid algorithms in advancing climate models, offering significant improvements in both performance and efficiency. This hybrid approach is highly transferable, enhancing feature selection and prediction capabilities for complex, multi-variable datasets. Similar models could be employed in areas like rainfall forecasting, air quality monitoring, and biodiversity predictions, where data dimensionality is a challenge.

The broader significance of this research extends far beyond its immediate application to global temperature prediction. By developing and validating a hybrid algorithm that combines the strengths of the Arithmetic Optimization Algorithm (AOA) and the Coati Optimization Algorithm (COA), this study provides a framework for addressing high-dimensional, complex problems in diverse domains. The demonstrated effectiveness of the hybrid AOA-COA in reducing dimensionality, avoiding overfitting, and improving computational efficiency highlights its potential as a versatile tool for predictive modeling in various fields.

One such field is biodiversity modeling, where predicting the distribution of species across landscapes requires handling large, complex datasets containing environmental, genetic, and ecological variables. The ability of the hybrid algorithm to select the most relevant features while maintaining high predictive accuracy can significantly enhance the modeling of species responses to climate change, habitat loss, and other anthropogenic pressures. This, in turn, could inform conservation strategies and policymaking aimed at preserving biodiversity in vulnerable ecosystems.

Another domain that stands to benefit is disaster management, where accurate predictions of extreme events like floods, droughts, or hurricanes are critical. The hybrid algorithm's capacity to process and analyze vast amounts of meteorological and environmental data can improve the accuracy and timeliness of disaster forecasts. This would enable better preparation, resource allocation, and response strategies, potentially saving lives and reducing economic losses. Furthermore, the algorithm's flexibility ensures that it can be tailored to specific challenges, such as optimizing evacuation routes or identifying high-risk areas for targeted interventions.

In addition to these applications, the methodology developed in this research has the potential to be generalized to other domains that rely on predictive analytics and optimization. Fields such as energy management, agricultural planning, and urban development often involve high-dimensional datasets where identifying key factors is essential for informed decision-making. By leveraging the principles of swarm intelligence and machine learning optimization demonstrated here, researchers and practitioners can adapt the hybrid AOA-COA framework to solve problems across a wide range of disciplines, amplifying its impact on real-world challenges.

This research thus not only advances the state of predictive modeling in climate science but also establishes a foundation for transformative applications in other critical domains, addressing some of the most pressing issues of our time.

6.2 Future Work

Applying machine learning (ML) techniques and swarm intelligence algorithms to predict global temperature and climate trends has shown promising results. However, there remain several avenues for future research and development that could enhance the accuracy and interpretability of these models. One significant area for future work lies in the comparison and also the integration of climate science models with machine learning approaches to develop more robust and comprehensive predictive models.

Future research could also explore additional hybrid swarm intelligence algorithms combining multiple optimization techniques to further improve feature selection and model accuracy. While our study demonstrated the effectiveness of a hybrid approach combining the Arithmetic optimization Algorithm (AOA) with the Coati Optimization Algorithm (COA), other combinations could be investigated. By testing different hybrid combinations, researchers can assess which algorithms are best suited for climate data, which is inherently high-dimensional. Additionally, exploring how these algorithms handle big data and real-time climate datasets could lead to advancements in predictive capabilities.

Moreover, ensemble modeling represents a promising avenue for future research. Future work could investigate the integration of ensemble techniques with the proposed hybrid swarm intelligence models. For example, utilizing a stacking approach, where predictions from multiple optimized models are combined using a meta-model, could enhance predictive power and robustness [Wolpert, 1992]. This line of inquiry could uncover novel ways to optimize feature selection and improve global temperature predictions, further advancing the field.

Bibliography

T. R. Karl, J. M. Melillo, and T. C. Peterson. *Global Climate Change Impacts in the United States*. Cambridge University Press, 2009.

IPCC. *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, Switzerland, 2023. doi: 10.59327/IPCC/AR6-9789291691647.

Intergovernmental Panel on Climate Change. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2021. URL https://www.ipcc.ch/report/ar6/wg1/.

K. Reddy. Impact of climate change on natural disasters. *Environmental Hazards Review*, 23(4):399–412, 2021.

Doruk Sen, Mehmet Fatih Huseyinoglu, and M. Erdem Günay. Prediction of global temperature anomaly by machine learning based techniques. *Neural Computing and Applications*, 35(21):15601–15614, 4 2023. doi: 10.1007/s00521-023-08580-3. URL https://doi.org/10.1007/s00521-023-08580-3.

Intergovernmental Panel on Climate Change. *Climate Change 2015: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2015. URL https://www.ipcc.ch/report/ar5/wg2/.

A. Kumar. Global warming, climate change and greenhouse gas mitigation. In A. Kumar, S. Ogita, and Y.Y. Yau, editors, *Biofuels: Greenhouse Gas Mitigation and Global Warming*. Springer, New Delhi, 2018. doi: 10.1007/978-81-322-3763-1_1.

Mark Holmstrom, Dylan Liu, and Christopher Vo. Machine learning applied to weather forecasting. Technical report, Stanford University, December 2016.

Neela Mishra, Hemant Kumar Soni, Sanjiv Sharma, and A. K. Upadhyay. Development and analysis of artificial neural network models for rainfall prediction by using time-series data. *MECS*, January 2018.

V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999. doi: 10.1109/72.788640.

H. Zheng. Analysis of global warming using machine learning. *Computational Water, Energy, and Environmental Engineering*, 7:127–141, 2018. doi: 10.4236/cweee.2018.73009.

M. Koc and A. Acar. Investigation of urban climates and built environment relations by using machine learning. *Urban Climate*, 37:100820, 2021.

X.Y. Chen and K.W. Chau. A hybrid double feedforward neural network for suspended sediment load estimation. *Water Resources Management*, 30:2179–2194, 2016. doi: 10.1007/s11269-016-1281-2.

T. Ladi, S. Jabalameli, and A. Sharifi. Applications of machine learning and deep learning methods for climate change mitigation and adaptation. *Environment and Planning B: Urban Analytics and City Science*, 49(4): 1314–1330, 2022. doi: 10.1177/23998083221085281.

Norman R. Draper and Harry Smith. *Applied Regression Analysis*. Wiley, New York, 3rd edition, 1998. ISBN 978-0-471-17082-2.

M. Abdel-Basset, L. Abdel-Fatah, and A. K. Sangaiah. Metaheuristic algorithms: A comprehensive review. In *Computational Intelligence for Multimedia Big Data on the Cloud With Engineering Applications*, pages 185–231. Elsevier, 2018. doi: 10.1016/B978-0-12-813314-9.00010-4.

T. Dokeroglu, E. Sevinc, T. Kucukyilmaz, and A. Cosar. A survey on new generation metaheuristic algorithms. *Computers & Industrial Engineering*, 137:106040, 2019. doi: 10.1016/j.cie.2019.106040.

Hongnian Zang, Shujun Zhang, and Kevin Hapeshi. A review of Nature-Inspired algorithms. *Journal of Bionic Engineering*, 7(S4):S232–S237, 10 2010. doi: 10.1016/s1672-6529(09)60240-7. URL https://doi.org/10.1016/s1672-6529(09)60240-7.

Parul Agarwal and Shikha Mehta. Nature-inspired algorithms: State-of-art, problems and prospects. *International Journal of Computer Applications*, 100(14):14–21, August 2014. doi: 10.5120/17593-8331. URL https://doi.org/10.5120/17593-8331.

P. K. Mandal and M. Thakur. Higher-order moments in portfolio selection problems: A comprehensive literature review. *Expert Systems With Applications*, 238:121625, 2024. doi: 10.1016/j.eswa.2023.121625.

J. F. Guerra, M. A. Llama, and V. Santibañez. A comparative study of swarm intelligence metaheuristics in ukf-based neural training applied to the identification and control of robotic manipulator. *Algorithms*, 16(8): 393, 2023. doi: 10.3390/a16080393.

H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):491–502, April 2005. doi: 10.1109/TKDE.2005.66.

A. Jović, K. Brkić, and N. Bogunović. A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1200–1205, Opatija, Croatia, 2015. doi: 10.1109/MIPRO.2015.7160458.

Mohamed A. Tawhid and Abdelmonem M. Ibrahim. Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm. *International Journal of Machine Learning and Cybernetics*, 11(3):573–602, 8 2019. doi: 10.1007/s13042-019-00996-5. URL https://doi.org/10.1007/s13042-019-00996-5.

L. Brezočnik, I. Fister, and V. Podgorelec. Swarm intelligence algorithms for feature selection: A review. *Applied Sciences*, 8(9):1521, 2018. doi: 10.3390/app8091521.

T. Vishvakarma. Climate change forecasting using machine learning algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 10(10):881–888, 2022. doi: 10.22214/ijraset.2022.47098.

M. Purushotham Reddy, A. Aneesh, K. Praneetha, and S. Vijay. Global warming analysis and prediction using data science. In *2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 1055–1059, Palladam, India, 2021. doi: 10.1109/I-SMAC52330.2021.9640944.

M. A. Tawhid and K. B. Dsouza. Hybrid binary bat enhanced particle swarm optimization algorithm for solving feature selection problems. *Applied Computing and Informatics*, 16(1/2):117–136, 2018. doi: 10.1016/j.aci.2018.04.001.

Essam Halim Houssein, Eman Saber, Yaser M. Wazery, and Abdelmgeid A. Ali. Swarm intelligence algorithms-based machine learning framework for medical diagnosis: A comprehensive review. In *Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems*, pages 45–68. Springer, 2022. doi: 10.1007/978-3-030-99079-4_4. URL https://link.springer.com/chapter/10.1007/978-3-030-99079-4_4.

Nebojsa Bacanin, Aleksandar Petrovic, Miodrag Zivkovic, Timea Bezdan, and Amit Chhabra. Enhanced salp swarm algorithm for feature selection. In *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation*, pages 123–138. Springer, 2021. doi: 10.1007/978-3-030-85626-7_57. URL https://link.springer.com/chapter/10.1007/978-3-030-85626-7_57.

Mohammad Dehghani, Zeinab Montazeri, Eva Trojovská, and Pavel Trojovský. Coati optimization algorithm: A new bio-inspired metaheuristic algorithm for solving optimization problems. *Knowledge-Based Systems*, 259:110011, 2023. ISSN 0950-7051. doi: https://doi.org/10.1016/

j.knosys.2022.110011. URL https://www.sciencedirect.com/science/article/pii/S0950705122011042.

L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. A. Al-qaness, and A. H. Gandomi. Arithmetic optimization algorithm: A new meta-heuristic algorithm for solving optimization problems. *Applied Mathematics and Computation*, 409:125577, 2021a. doi: 10.1016/j.amc.2021.125577.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.

James Hansen, Reto Ruedy, Makiko Sato, and Ken Lo. Global surface temperature change. *Reviews of Geophysics*, 48(4):RG4004, 2010. doi: 10.1029/2010RG000345.

Kaggle. Climate change dataset. https://www.kaggle.com/datasets/econdata/climate-change, 2019.

T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, and P. M. Midgley, editors. *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 2013.

V. Ramaswamy, O. Boucher, J. Haigh, D. Hauglustaine, J. Haywood, G. Myhre, and F. Stordal. Radiative forcing of climate change. In *Climate Change 2001: The Scientific Basis*. Cambridge University Press, 2001.

Hannah Ritchie, Pablo Rosado, and Max Roser. Co2 and greenhouse gas emissions. *Our World in Data*, 2023. https://ourworldindata.org/co2-and-greenhouse-gas-emissions.

Hakan Saritas. Greenhouse gases dataset. https://www.kaggle.com/datasets/hakansaritas/greenhouse-gases, 2022.

Daniel Dias. Co2 and greenhouse gas emissions dataset. https://www.kaggle.com/datasets/danielrpdias/co2-and-greenhouse-gas-emissions/data, 2023.

Climate change indicators: U.S. and global temperature — US EPA, 6 2024. URL https://www.epa.gov/climate-indicators/climate-change-indicators-us-and-global-temperature.

Rob J. Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, 2nd edition, 2018. Available at https://otexts.com/fpp2/.

R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

M. Kuhn and K. Johnson. *Applied Predictive Modeling*. Springer, 2013.

G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning*. Springer, 2013.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

D. C. Montgomery, E. A. Peck, and G. G. Vining. *Introduction to Linear Regression Analysis*, volume 821. John Wiley & Sons, 2012.

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992. doi: 10.1080/00031305.1992.10475879.

T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, August 2016. doi: 10.1145/2939672.2939785.

X. Zhang, C. Yan, C. Gao, et al. Predicting missing values in medical data via xgboost regression. *Journal of Healthcare Informatics Research*, 4:383–394, 2020. doi: 10.1007/s41666-020-00077-1.

L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-qaness, and A. H. Gandomi. Aquila optimizer: A novel meta-heuristic optimization algorithm. *Computers & Industrial Engineering*, 157:107250, 2021b. doi: 10.1016/j.cie.2021.107250.

P. Trojovský and M. Dehghani. Pelican optimization algorithm: A novel nature-inspired algorithm for engineering applications. *Sensors*, 22(3):855, 2022. doi: 10.3390/s22030855.

C. Blum and X. Li. Swarm intelligence in optimization. In *Swarm Intelligence*, pages 43–85. Springer, Berlin/Heidelberg, Germany, 2008.

M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY, USA, 1979.

J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, USA, 1975.

S. Saremi, S. Mirjalili, and A. Lewis. Grasshopper optimisation algorithm: Theory and application. *Advances in Engineering Software*, 105:30–47, 2017. doi: 10.1016/j.advengsoft.2017.01.004.

E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi. Gsa: A gravitational search algorithm. *Information Sciences*, 179(13):2232–2248, 2009. doi: 10.1016/j.ins.2009.03.004.

A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen. Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, 97:849–872, 2019. doi: 10.1016/j.future.2019.02.028.

S. Mirjalili, S. M. Mirjalili, and A. Lewis. Grey wolf optimizer. *Advances in Engineering Software*, 69:46–61, 2014. doi: 10.1016/j.advengsoft.2013.12.007.

S. Mirjalili and A. Lewis. The whale optimization algorithm. *Advances in Engineering Software*, 95:51–67, 2016. doi: 10.1016/j.advengsoft.2016.01.008.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995. doi: 10.1109/ICNN.1995.488968.

Jiankai Xue and Bo Shen. Dung beetle optimizer: a new meta-heuristic algorithm for global optimization. *The Journal of Supercomputing*, 79 (7):7305–7336, 11 2022. doi: 10.1007/s11227-022-04959-6. URL https://doi.org/10.1007/s11227-022-04959-6.

A. F. Siegel. Robust regression using repeated medians. *Biometrika*, 69(1): 242–244, 1982.

J. M. Wooldridge. *Introductory Econometrics: A Modern Approach*. Cengage Learning, 6th edition, 2015.

G. Chandrashekar and F. Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014. doi: 10.1016/j.compeleceng.2013.11.024.

I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, 46:389–422, 2002. doi: 10.1023/A:1012487302797.

T. Ishitaki, T. Oda, and L. Barolli. A neural network based user identification for tor networks: Data analysis using friedman test. In *2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pages 7–13, Crans-Montana, Switzerland, 2016. doi: 10.1109/WAINA.2016.143.

J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

M. Hollander, D. A. Wolfe, and E. Chicken. *Nonparametric Statistical Methods*. John Wiley & Sons, 3rd edition, 2013.

R. L. Iman and J. M. Davenport. Approximations of the critical region of the friedman statistic. *Communications in Statistics-Theory and Methods*, 9(6):571–595, 1980.

David J. Sheskin. Handbook of Parametric and Nonparametric Statistical Procedures. *Technometrics*, 46(3):369–370, 8 2004. doi: 10.1198/tech.2004.s209. URL https://doi.org/10.1198/tech.2004.s209.

Julián Derrac, Salvador García, Daniel Molina, and Francisco Herrera. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, 1(1):3–18, March 2011. doi: 10.1016/j.swevo.2011.02.002.

El-Ghazali Talbi. *Metaheuristics: From Design to Implementation*. John Wiley & Sons, Hoboken, NJ, USA, 2009.

Xin-She Yang and Suash Deb. Cuckoo search via lévy flights. In *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pages 210–214. IEEE, 2009. doi: 10.1109/NABIC.2009.5393690.

H. Damon Matthews, Nathan P. Gillett, Peter A. Stott, and Kirsten Zickfeld. The proportionality of global warming to cumulative carbon emissions. *Nature*, 459(7248):829–832, 2009.

H. Damon Matthews, Susan Solomon, and Raymond Pierrehumbert. Cumulative carbon as a policy framework for achieving climate stabilization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1974):4365–4379, 2012.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566 (7743):195–204, 2019.

David Rolnick, Priya L. Donti, Lynn H. Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Alán Aspuru-Guzik Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433*, 2019.

David A. Randall, Richard A. Wood, Sandrine Bony, et al. Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, 2007.

M. Hashemitaheri, S. M. R. Mekarthy, and H. P. Cherukuri. Prediction of specific cutting forces and maximum tool temperatures in orthogonal machining by support vector and gaussian process regression methods. *Procedia Manufacturing*, 2020. doi: 10.1016/j.promfg.2020.05.139.

H. Liu, Q. Liu, M. Cai, K. Chen, L. Ma, W. Meng, Z. Zhou, and Q. Ai. Attention-based multi-semantic dynamical graph convolutional network for eeg-based fatigue detection. *Frontiers in Neuroscience*, 2023. doi: 10.3389/fnins.2023.00000.

X. S. Yang. *Nature-Inspired Metaheuristic Algorithms.* Luniver Press, 2010.

E. Aarts and J. K. Lenstra. *Local Search in Combinatorial Optimization.* Princeton University Press, 1997.

H. H. Hoos and T. Stützle. *Stochastic Local Search: Foundations & Applications*. Elsevier, 2005.

R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1137–1143, 1995.

David H Wolpert. Stacked generalization. *Neural Networks*, 5(2):241–259, 1992.